



# STIC Search Report

## Biotech-Chem Library

STIC Database Tracking Number: 98584

TO: Mark Navarro  
Location: 8A15  
Art Unit: 1645  
Friday, July 11, 2003

Case Serial Number: 10/091442

From: Barb O'Bryen  
Location: Biotech-Chem Library  
CM1-6A05  
Phone: 308-4291 *BOB*

barbara.obryen@uspto.gov

### Search Notes

O'Bryen, Barbara

From: Navarro, Albert  
Sent: Thursday, July 10, 2003 12:47 PM  
To: O'Bryen, Barbara  
Subject: 10/091,442

Mark Navarro  
1645  
306-3225  
8A15

Hello Barb,

Could I trouble you for a search and an interference search of:

SEQ ID NO: 1-9 and 34

Thanks

Mark

7/14/93  
(7/17/92 FP)

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:44:31 ; Search time 8.29522 Seconds  
(without alignments)  
471.946 Million cell updates/sec

Title: US-10-091-442-3  
Perfect score: 100  
Sequence: 1 VERVDFTNHLEDFRRNINK 19

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 671580 seqs, 206047115 residues  
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SPREMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query	Length	DB ID	Description
1	71.0	380	11	Q9D695	Q9d695 mus musculus
2	68.0	380	11	Q920J5	Q920J5 rattus norv
3	51.5	509	12	Q9YQ03	Q9YQ03 myxoma viru
4	51.5	509	12	Q9PX36	Q9PX36 myxoma viru
5	50.0	338	4	Q9BYR8	Q9BYR8 homo sapien
6	50.0	2020	5	Q8T2M1	Q8T2M1 dictyosteli
7	48.0	380	5	Q8WQX0	Q8WQX0 rhinipicephal
8	48.0	472	5	Q09657	Q09657 caenorhabdi
9	47.0	245	16	Q8XM42	Q8XM42 clostotridium
10	47.0	419	4	Q43611	Q43611 homo sapien
11	47.0	672	5	Q25586	Q25586 onchocerca
12	47.0	1024	4	Q14527	Q14527 homo sapien
13	47.0	1046	4	Q9UDT6	Q9UDT6 homo sapien
14	47.0	1046	11	Q920H8	Q920H8 mus musculu
15	47.0	1046	11	Q55156	Q55156 rattus norv
16	47.0	1047	11	Q9EP81	Q9EP81 mus musculu

17	46.0	173	10	Q946D0	Q946D0 theobroma c
18	46.0	174	10	Q945Y0	Q945Y0 theobroma c
19	46.0	231	4	Q8WM89	Q8WM89 homo sapien
20	46.0	369	4	Q9BYF7	Q9BYF7 homo sapien
21	46.0	427	5	Q9U114	Q9U114 drosophila
22	46.0	427	5	Q9VFC2	Q9VFC2 drosophila
23	45.0	309	16	Q92PT5	Q92PT5 rhizobium m
24	45.0	361	4	Q9BX29	Q9BX29 homo sapien
25	45.0	443	5	Q76878	Q76878 drosophila
26	45.0	539	11	Q9DCD5	Q9DCD5 mus musculu
27	45.0	673	16	Q24911	Q24911 helicobacte
28	45.0	673	16	Q92MW7	Q92MW7 helicobacte
29	44.5	330	16	Q97FEL	Q97FEL clostotridium
30	44.0	311	2	Q9WRC2	Q9WRC2 escherichia
31	44.0	348	16	Q9PR45	Q9PR45 ureaplasma
32	44.0	369	10	Q9L790	Q9L790 arabidopsis
33	44.0	542	10	Q9FY07	Q9FY07 capsicum an
34	44.0	617	2	Q32544	Q32544 escherichia
35	44.0	709	5	Q9GRG1	Q9GRG1 tetrahymena
36	43.5	119	16	Q55617	Q55617 synecocyst
37	43.0	58	16	Q97Q04	Q97Q04 streptococ
38	43.0	173	11	Q08882	Q08882 rattus norv
39	43.0	175	5	Q9VXT8	Q9VXT8 drosophila
40	43.0	259	16	Q9RQW0	Q9RQW0 neisseria m
41	43.0	301	5	Q95PD8	Q95PD8 calliphora
42	43.0	301	16	Q25565	Q25565 helicobacte
43	43.0	363	16	Q92KU2	Q92KU2 helicobacte
44	43.0	388	16	Q8XN50	Q8XN50 clostotridium
45	43.0	468	10	Q9S1B7	Q9S1B7 arabidopsis

## ALIGNMENTS

## RESULT 1

ID	Q9D695	PRELIMINARY:	PRT:	380 AA.
AC	Q9D695:			
DT	01-JUN-2001 (TREMBLrel. 17, Created)			
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE	4631416M05R1K protein (Megsin).			
GN	SERPINB7 OR 4631416M05R1K.			
OS	Mus musculus (Mouse).			
CC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=SKIN;			
RX	MEDLINE=21085660; PubMed=11217851;			
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,			
RA	Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,			
RA	Alzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,			
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,			
RA	Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,			
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,			
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,			
RA	Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,			
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,			
RA	Blake J., Boffelli D., Bojunga N., Carlini P., de Bonaudo M.F.,			
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,			
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,			
RA	Lyon P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,			
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,			
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,			
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Winking L.,			
RA	Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,			
RA	Hayashizaki Y.,			
RT	"Functional annotation of a full-length mouse cDNA collection.";			
RL	Nature 409:685-690(2001).			
RN	[2]			
RP	SEQUENCE FROM N.A.			

```

RX MEDLINE-21368006; PubMed-11473647;
RA Nengaku M., Miyata T., Suzuki D., Umezono T., Hashimoto T., Wada T.,
RA Yagi M., Nagano N., Inagi R., Kurokawa K.;
RT "Cloning of rodent megilin revealed its up-regulation in
RT mesangio proliferative nephritis.";
RL Kidney Int. 60:641-652(2001).
DR EMBL: AK014524; BAB29410.1; -
DR HSSP: P05121; 1A7C.
DR MGD: MGI:2151053; Serpinb7.
DR InterPro: IPR000215; Serpin.
DR Pfam: PF00079; Serpin.
DR SMART: SM00093; SERPIN.1.
DR PROSITE: PS00284; SERPIN; UNKNOWN_1.
SQ SEQUENCE 380 AA; 43050 MW; C9240272BCEP9CF4 CRC64;

Query Match 71.0%; Score 71; DB 11; Length 380;
Best Local Similarity 73.7%; Pred. No. 0.0027;
Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 VERVDFTNHLDTRRNIN 19
Db 126 VERVDFTNDYDTRFKIN 144

RESULT 2
Q920J5 PRELIMINARY; PRT; 380 AA.
AC 0920J5;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE Megsin.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-21368006; PubMed-11473647;
RA Nengaku M., Miyata T., Suzuki D., Umezono T., Hashimoto T., Wada T.,
RA Yagi M., Nagano N., Inagi R., Kurokawa K.;
RT "Cloning of rodent megilin revealed its up-regulation in
RT mesangio proliferative nephritis.";
RL Kidney Int. 60:641-652(2001).
CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.
DR EMBL: AF105329; AAL16769.1; -
DR InterPro: IPR000215; Serpin.
DR Pfam: PF00079; Serpin.1.
DR PROSITE: PS00284; SERPIN; UNKNOWN_1.
KW Serpin.
SQ SEQUENCE 380 AA; 42821 MW; D8076CABEE2C2FBC CRC64;

Query Match 68.0%; Score 68; DB 11; Length 380;
Best Local Similarity 68.4%; Pred. No. 0.0081;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 VERVDFTNHLDTRRNIN 19
Db 126 VERVDFTNDYDTRFKIN 144

RESULT 3
Q9Y003 PRELIMINARY; PRT; 509 AA.
AC 09Y003;
DT 01-MAY-1999 (TREMblrel. 10, Created)
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE M-T6 protein.
GN M-T6.
OS Myxoma virus (strain Lausanne).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;

```

```

OC Leporipoxvirus.
OX NCBI_TaxID=31530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LAUSANNE;
RA Salim S., Meissner K., Neubert A., Doehner L., Becher D.;
RT "DNA sequence of the M-T6 gene of Myxoma virus strain Lausanne.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ012283; CA009975.1; -
DR InterPro: IPR000210; BTB_POZ.
DR InterPro: IPR000875; Cectropin.
DR InterPro: IPR001798; Kelch.
DR Pfam: PF00651; BTB; 1.
DR Pfam: PF01344; Kelch; 5.
DR SMART: SM00225; BTB; 1.
DR PROSITE: PS50097; BTB; 1.
DR PROSITE: PS00268; CECTROPIN; UNKNOWN_1.
SQ SEQUENCE 509 AA; 58166 MW; 5A7799B6015C6BFA CRC64;

Query Match 51.5%; Score 51.5; DB 12; Length 509;
Best Local Similarity 41.4%; Pred. No. 5.2;
Matches 12; Conservative 3; Mismatches 3; Indels 11; Gaps 1;

QY 1 VERVDFTN-----HLEDTRRNIN 18
Db 109 VSRVDPTNCVSARFQFAETVHIEDLKRILN 137

RESULT 4
Q9PX36 PRELIMINARY; PRT; 509 AA.
AC 09PX36;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE M006L protein.
GN M006L OR M006R.
OS Myxoma virus (strain Lausanne).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Leporipoxvirus.
OX NCBI_TaxID=31530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LAUSANNE;
RX MEDLINE-20032073; PubMed-10562494;
RA Cameron C., Hota-Mitchell S., Chen L., Barrett J., Cao J.X.,
RA Macaulay C., Miller D., Evans D., McFadden G.;
RT "The complete DNA sequence of myxoma virus.";
RL Virology 264:298-318(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-LAUSANNE;
RA Cameron C., Hota-Mitchell S., Chen L., Barrett J., Cao J.X.,
RA Macaulay C., Miller D., Evans D., McFadden G.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF170726; AAF15052.1; -
DR EMBL: AF170726; AAF14893.1; -
DR InterPro: IPR000210; BTB_POZ.
DR InterPro: IPR000875; Cectropin.
DR InterPro: IPR001798; Kelch.
DR Pfam: PF00651; BTB; 1.
DR Pfam: PF01344; Kelch; 5.
DR SMART: SM00225; BTB; 1.
DR PROSITE: PS50097; BTB; 1.
DR PROSITE: PS00268; CECTROPIN; UNKNOWN_1.
SQ SEQUENCE 509 AA; 58166 MW; 5D0129AA03DE4B7D CRC64;

Query Match 51.5%; Score 51.5; DB 12; Length 509;
Best Local Similarity 41.4%; Pred. No. 5.2;
Matches 12; Conservative 3; Mismatches 3; Indels 11; Gaps 1;

QY 1 VERVDFTN-----HLEDTRRNIN 18
Db 109 VSRVDPTNCVSARFQFAETVHIEDLKRILN 137

```

DB 109 VSRVDPNTNCVSAFOFAETVHIEDIKRNLN 137

## RESULT 5

Q9BYF8 PRELIMINARY; PRT; 338 AA.  
AC Q9BYF8; 01-JUN-2001 (TReMBLrel. 17, Created)  
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)  
DE 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
SCSALB.  
GN SCSALB.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Sumitani Y., Kishi F., Murakami A., Sakaguchi Y., Kato H.;  
RT "Novel Forms of SCC Antigen Transcripts Produced by Alternative  
RT Splicing";  
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.  
DR HSSP; P01008; IATH.  
DR EMBL; AB046399; BAB40772.1; -  
DR InterPro: IPR000215; Serpin.  
DR Pfam: PF00079; serpin.1.  
DR SMART: SM00093; SERPIN.1.  
DR PROSITE: PS00284; SERPIN; 1.  
KW Serpin.  
SQ SEQUENCE 338 AA; 38519 MW; 5AB9CEFA80C1ABDF CRC64;

Query Match 50.0%; Score 50; DB 4; Length 338;  
Best Local Similarity 55.6%; Pred. No. 5.8;  
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 VERVDFTNHLEDTNRNIN 18

DB 131 VESYDFANAPESKKNIN 148

## RESULT 6

Q8T2M1 PRELIMINARY; PRT; 2020 AA.  
AC Q8T2M1; 01-JUN-2002 (TReMBLrel. 21, Created)  
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)  
DE Hypothetical 232.6 kDa protein.  
OS Dictyostelium discoideum (Slime mold).  
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.  
OX NCBI\_TaxID=44685;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Gloeckner G., Eichinger L., Szafranski K., Pachepat J., Dear P.,  
RA Lehmann R., Baumgart C., Parra G., April J.F., Gilgo R., Kumpf K.,  
RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;  
RT "Sequence and Analysis of Chromosome 2 of Dictyostellium";  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC115594; AAL92302.1; -  
KW Hypothetical protein.  
SQ SEQUENCE 2020 AA; 232615 MW; 01E221CE63114DD3 CRC64;

Query Match 50.0%; Score 50; DB 5; Length 2020;  
Best Local Similarity 56.2%; Pred. No. 42;  
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 4 VDFTNHLEDTNRNIN 19

DB 967 IDDSNHSEDTTNSINK 982

## RESULT 7

Q8WQX0 PRELIMINARY; PRT; 380 AA.  
AC Q8WQX0; 01-MAR-2002 (TReMBLrel. 20, Created)  
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)  
DE 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
OS Rhiniphalus appendiculatus (Brown ear tick).  
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;  
Parasitiformes; Ixodida; Ixodidae; Rhiniphalus.  
OX NCBI\_TaxID=34631;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Milenga A., Sugimoco C., Onuma M.;  
RT "Blood meal feeding-induced serine proteinase inhibitor-2 from  
RT Rhiniphalus appendiculatus";  
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY035780; AAK61376.1; -  
DR InterPro: IPR000215; Serpin.  
DR Pfam: PF00079; serpin.1.  
DR SMART: SM00093; SERPIN.1.  
DR PROSITE: PS00284; SERPIN; 1.  
SQ SEQUENCE 380 AA; 42729 MW; 762B7C44E1F611F2 CRC64;

Query Match 48.0%; Score 48; DB 5; Length 380;  
Best Local Similarity 44.4%; Pred. No. 14;  
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 VERVDFTNHLEDTNRNIN 18

DB 120 IESYDFKTHYKVRQOVN 137

## RESULT 8

Q09657 PRELIMINARY; PRT; 472 AA.  
AC Q09657; 01-JAN-1999 (TReMBLrel. 09, Created)  
DT 01-JAN-1999 (TReMBLrel. 09, Last sequence update)  
DE 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
DE Hypothetical 51.4 kDa protein ZK1320.9 in chromosome II.  
GN ZK1320.9.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
OC Rhabditidae; Pelodermidae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN-BRISTOL N2;  
RA Berks M.;  
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: STRONG, TO C.KLUVVERI ORFZ.  
DR EMBL; Z46934; CA87047.1; -  
DR WormPep: ZK1320.9; CE01707.  
DR InterPro: IPR003702; AcetylCoA\_hydro.  
DR InterPro: IPR001220; Lectin\_legh.  
DR Pfam: PF02550; AcetylCoA\_hydro.1.  
DR PROSITE: PS00307; LECTIN\_LEGHME\_BETA; UNKNOWN.1.  
KW Hypothetical protein.  
SQ SEQUENCE 472 AA; 51369 MW; 08C3351E11B7C348 CRC64;

Query Match 48.0%; Score 48; DB 5; Length 472;  
Best Local Similarity 66.7%; Pred. No. 18;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 DFTNHLEDTNRNIN 16

DB 328 DFTNHLEDTNRNIN 339

## RESULT 9

Q8XM42

ID 08XM42 PRELIMINARY; PRT; 245 AA.  
 AC 08XM42;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Hypothetical protein CPE0848.  
 GN CPE0848.  
 OS Clostridium perfringens.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;  
 OC Clostridiales; Clostridiaceae; Clostridium.  
 RX NCBI\_TaxID=1502;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=13 / TYPE A;  
 RX PubMed=11792842;  
 RA Shimizu T., Ohtsuka K., Hirakawa H., Ohshima K., Yamashita A.,  
 RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.,  
 RT "Complete genome sequence of Clostridium perfringens, an anaerobic  
 RT flesh-eater.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).  
 DR EMBL: AP003188; BAB80554.1;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 245 AA; 28398 MW; 229703EAC24770D2 CRC64;

Query Match 47.0%; Score 47; DB 16; Length 245;  
 Best Local Similarity 57.1%; Pred. No. 12;  
 Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 5 DFTNHLDETRRNIN 18  
 Db 113 DLSNLSIENRRNIN 126

RESULT 10  
 ID 043611 PRELIMINARY; PRT; 419 AA.  
 AC 043611;  
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Probable CLIP protein (Fragment).  
 GN WSCR4.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96411691; PubMed=8812460;  
 RA Osborne L.R., Martindale D.W., Scherer S.W., Shi X.-M., Huizenga J.,  
 RA Heng H.H.Q., Costa T., Pober B., Lew L., Brinkman J., Rommens J.,  
 RA Koop B.F., Tsui L.-C.;  
 RT "Identification of genes from a 500-kb region at 7q11.23 that is  
 RT commonly deleted in Williams syndrome patients.";  
 RL Genomics 36:328-336(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Martindale D.W., Koop B.F.;  
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF041059; AAB96784.1;  
 DR EMBL: AF041055; AAB96784.1; JOINED.  
 DR EMBL: AF041056; AAB96784.1; JOINED.  
 DR EMBL: AF041057; AAB96784.1; JOINED.  
 DR EMBL: AF041058; AAB96784.1; JOINED.  
 DR InterPro: IPR000938; CAP-Gly.  
 DR Pfam: PF01302; CAP\_GLY; 2.  
 DR PROSITE: PS00845; CAP\_GLY\_1; UNKNOWN\_2.  
 DR NON\_TER 1  
 FT NON\_TER 1  
 SQ SEQUENCE 419 AA; 44825 MW; 547CA46012255C25 CRC64;

Query Match 47.0%; Score 47; DB 4; Length 419;  
 Best Local Similarity 50.0%; Pred. No. 22;

Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
 OY 2 ERVDFTNHLDETRRNIN 17  
 Db 386 ERVDLSNQLFEERRKV 401

RESULT 11  
 ID 025586 PRELIMINARY; PRT; 672 AA.  
 AC 025586;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE Paramyosin-related protein (Fragment).  
 OS Onchocerca gibsoni.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;  
 OC Onchocercidae; Onchocerca.  
 RX NCBI\_TaxID=6284;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96360163; PubMed=8719949;  
 RA Zhang D., Miller D.J.;  
 RT "Characterization of a novel non-muscle myosin-related protein from  
 RT Onchocerca gibsoni.";  
 RL Int. J. Parasitol. 25:1385-1391(1995).  
 DR EMBL: U20609; AAB50272.1;  
 DR InterPro: IPR000727; T-SNARE.  
 DR NON\_TER 1  
 FT NON\_TER 1  
 SQ SEQUENCE 672 AA; 80074 MW; C610249CEFOFFED1 CRC64;

Query Match 47.0%; Score 47; DB 5; Length 672;  
 Best Local Similarity 60.0%; Pred. No. 38;  
 Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 VERVDFTNHLDETRR 15  
 Db 83 ISVDVDFANQLEDERR 97

RESULT 12  
 ID 014527 PRELIMINARY; PRT; 1024 AA.  
 AC 014527;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE KIAA0291 protein (Fragment).  
 GN KIAA0291.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRIN;  
 RA Ohara O., Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N.,  
 RA Nomura N.;  
 RT "Prediction of the coding sequences of unidentified human genes.";  
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB006629; BAA2260.2;  
 DR InterPro: IPR000938; CAP-Gly.  
 DR Pfam: PF01302; CAP\_GLY; 2.  
 DR PROSITE: PS00845; CAP\_GLY\_1; UNKNOWN\_2.  
 DR NON\_TER 1  
 FT NON\_TER 1  
 SQ SEQUENCE 1024 AA; 112981 MW; 12278FC36788B620 CRC64;

Query Match 47.0%; Score 47; DB 4; Length 1024;  
 Best Local Similarity 50.0%; Pred. No. 60;  
 Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 2 ERVDFTNHLDETRRNIN 17

```

Db          440 EKVDLSNLEEBERRKV 455

RESULT 13
G9UDT6      PRELIMINARY;      PRT: 1046 AA.
ID 09UDT6
AC 09UDT6
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE KIA0291, similar rodent cytoplasmic linker protein CLIP-115 and
  reslin.
GN WUSC:H.GS166C05.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=99063792; PubMed=9847074;
RT "Toward a complete human genome sequence."
RL Genome Res. 8:1097-1108(1998).
RN 12
RP SEQUENCE FROM N.A.
RA Bemis G., Langston Y., Kellen J.;
RT "The sequence of Homo sapiens BAC clone GS1-166C5."
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN 13
RP SEQUENCE FROM N.A.
RA Waterston R.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC005015; AAF03524.1;
DR InterPro: IPR000938; CAP_Gly; 2.
DR Pfam: PF01302; CAP_Gly; 2.
DR PROSITE: PS00845; CAP_Gly_1; UNKNOWN_2.
SQ SEQUENCE 1046 AA; 115837 MW; 7A0B3C796E1C6E25 CRC64;

Query Match          47.0%; Score 47; DB 4; Length 1046;
Best Local Similarity 50.0%; Pred. No. 61;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Oy          2 ERVDFTNHLDETRRNI 17
          1:11:11:11:11:
Db          427 EKVDLSNLEEBERRKV 442

RESULT 14
G9ZOH8      PRELIMINARY;      PRT: 1046 AA.
ID 09ZOH8
AC 09ZOH8
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE CLIP-115.
GN CYN2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 11
RP SEQUENCE FROM N.A.
RX STRAIN=129;
RX MEDLINE=99017971; PubMed=9799601;
RA Hoogenraad C.C., Eussen B.H.J., Langeveld A., van Haperen R.,
  Witterberg S., Mouters C.H., Grosveld F., de Zeeuw C.I., Galjart N.;
RT "The murine CYN2 gene: genomic organization, chromosome localization,
  and comparison to the human gene that is located within the 7q11.23
  Williams syndrome critical region."
RL Genomics 53:348-358(1998).
DR EMBL: AJ228865; CAA13069.1;
DR EMBL: AJ228866; CAA13069.1; JOINED.

```

```

DR EMBL: AJ228867; CAA13069.1; JOINED.
DR EMBL: AJ228868; CAA13069.1; JOINED.
DR EMBL: AJ228869; CAA13069.1; JOINED.
DR EMBL: AJ228870; CAA13069.1; JOINED.
DR EMBL: AJ228871; CAA13069.1; JOINED.
DR EMBL: AJ228872; CAA13069.1; JOINED.
DR EMBL: AJ228873; CAA13069.1; JOINED.
DR EMBL: AJ228874; CAA13069.1; JOINED.
DR EMBL: AJ228875; CAA13069.1; JOINED.
DR EMBL: AJ228876; CAA13069.1; JOINED.
DR EMBL: AJ228877; CAA13069.1; JOINED.
DR EMBL: AJ228878; CAA13069.1; JOINED.
DR EMBL: AJ228879; CAA13069.1; JOINED.
DR EMBL: AJ228880; CAA13069.1; JOINED.
DR EMBL: AJ228883; CAA13069.1;
DR MGD: MGI:131316; CYN2.
DR InterPro: IPR000938; CAP_Gly.
DR InterPro: IPR000812; TFIIB_euk.
DR Pfam: PF01302; CAP_Gly; 2.
DR PROSITE: PS00845; CAP_Gly_1; UNKNOWN_1.
DR PROSITE: PS00782; TFIIB; UNKNOWN_1.
SQ SEQUENCE 1046 AA; 115494 MW; A2B9C28FC95B808C CRC64;

Query Match          47.0%; Score 47; DB 11; Length 1046;
Best Local Similarity 50.0%; Pred. No. 61;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Oy          2 ERVDFTNHLDETRRNI 17
          1:11:11:11:11:
Db          428 EKVDLSNLEEBERRKV 443

RESULT 15
G551S6      PRELIMINARY;      PRT: 1046 AA.
ID 0551S6
AC 0551S6
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE CLIP-115 protein.
GN CLIP-115.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=98087115; PubMed=9427243;
RA de Zeeuw C.I., Hoogenraad C.C., Goedknegt E., Hertzberg E.,
  Neubauer A., Grosveld F.G., Galjart N.J.;
RT "CLIP-115, a novel brain specific cytoplasmic linker protein, mediates
  the localisation of dendritic lamellar bodies."
RL Neuron 19:1187-1199(1997).
DR EMBL: AJ000485; CAA04123.1;
DR InterPro: IPR000938; CAP_Gly.
DR InterPro: IPR000812; TFIIB_euk.
DR Pfam: PF01302; CAP_Gly; 2.
DR PROSITE: PS00845; CAP_Gly_1; UNKNOWN_2.
DR PROSITE: PS00782; TFIIB; UNKNOWN_1.
SQ SEQUENCE 1046 AA; 115480 MW; 72E6CE9F76D2A1D0 CRC64;

Query Match          47.0%; Score 47; DB 11; Length 1046;
Best Local Similarity 50.0%; Pred. No. 61;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Oy          2 ERVDFTNHLDETRRNI 17
          1:11:11:11:11:
Db          428 EKVDLSNLEEBERRKV 443

```

Search completed: July 11, 2003, 11:58:54  
Job time : 11.2952 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:42:16 ; Search time 9.75468 Seconds  
(without alignments)  
232.223 Million cell updates/sec

Title: US-10-091-442-4

Sequence: 1 SYIEVTEEGTATATG 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Total number of hits satisfying chosen parameters: 908470

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Maximum DB seq length: 200000000000

Post-processing: Minimum match 0%

Listing first 45 summaries

Database : A\_Geneseq\_101002:\*

[illegible]

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	83	100.0	380	15	AAR48379	Hunan megakaryocy
2	83	100.0	380	20	AAV08854	Hunan megaslin prote
3	83	100.0	380	21	AAB24142	Hunan megaslin prote
4	83	100.0	380	22	AAB83075	Hunan megaslin prote
5	69	83.1	17	15	AAR57110	Hunan megakaryocy
6	69	83.1	217	20	AAV29329	Hunan secreted pro
7	69	83.1	217	22	AAU039055	Hunan secreted pro
8	69	83.1	217	23	ABB55764	Hunan polypeptide
9	69	83.1	340	22	AAU16828	Hunan novel secre
10	69	83.1	368	20	AAV08256	Mouse megaslin prote

11	69	83.1	368	21	AAB24151
12	69	83.1	368	22	AAB83077
13	69	83.1	379	21	AAV15155
14	69	83.1	617	21	AAV15156
15	67	80.7	135	22	AAV72661
16	67	80.7	316	20	AAV28644
17	67	80.7	391	20	AAV41029
18	67	80.7	391	20	AAV25926
19	67	80.7	391	21	AAAB11314
20	67	80.7	391	21	AAAB10953
21	67	80.7	391	21	AAAB10954
22	67	80.7	391	22	AAV72651
23	67	80.7	391	23	ABAB74946
24	67	80.7	400	20	AAV41030
25	67	80.7	400	21	AAAB11315
26	67	80.7	400	22	AAV72652
27	67	80.7	400	23	ABAB74947
28	67	80.7	1120	22	ABG28279
29	66	79.5	392	22	ABBB2205
30	66	79.5	424	22	ABBB2205
31	66	79.5	431	22	ABBB64428
32	65	78.3	380	20	AAV08255
33	65	78.3	380	21	AAAB24150
34	65	78.3	380	22	AAAG64286
35	65	78.3	380	22	AAAB83076
36	62	74.7	116	22	AAU02921
37	62	74.7	188	22	AAU02923
38	62	74.7	372	22	AAU02963
39	62	74.7	374	22	ABBB6309
40	62	74.7	376	22	AAAB58176
41	62	74.7	376	22	AAAB11125
42	62	74.7	379	16	AAAB64159
43	62	74.7	379	17	AAAB9367
44	62	74.7	389	22	AAU02973
45	62	74.7	400	22	AAU02962

## ALIGNMENTS

RESULT 1	
AA048379	
ID	AA048379 standard; Protein; 380 AA.
XX	
AC	AA048379;
XX	
DT	16-AUG-1994 (first entry)
XX	
DE	Human megakaryocyte differentiation factor.
XX	
KW	Human megakaryocyte differentiation factor; MDF; thrombopoietin
KW	haematopoietic stimulating factor; thrombocytopoiesis; platelet;
KW	bone marrow transplantation; cancer chemotherapy.
XX	
OS	Homo sapiens.
PN	EP583864-A.
XX	
PD	23-FEB-1994.
XX	
PF	19-JUL-1993; 93EP-0305654.
XX	
PR	17-JUL-1992; 92JP-0212305.
XX	
PR	04-MAR-1993; 93JP-0067339.
XX	
PA	(SUNR ) SUNTORY LTD.
PA	(TSUJ/) TSUJIMOTO M.
PI	Ishida N, Iwasa F, Kurihara T, Miura K, Nakazato H;
XX	
PI	Tsujimoto M, Tsunoka N, Yamaguchi N, Yamachi K;
XX	
WP:	1994-058782/08.
DR	N-PSDB; AA056670.

```

XX New megakaryocyte differentiation factor - isolated from human
PT epidermoid carcinoma cells, used to treat conditions involving a
PT decrease in platelets
XX
PS Claim 7; Page 30-32; 47pp; English.
XX
CC Human MDF can be isolated from a culture of human epidermoid
CC carcinoma A431 cells in protein-free medium. The MDF stimulates
CC differentiation of megakaryocytes from myeloid cells in the presence
CC of IL-3. The MDF acts in vivo as a thrombopoietin making it useful
CC for treatment of diseases involving a decrease in platelet number
CC (esp. thrombocytopenia) such as occurs in bone marrow
CC transplantation and in chemotherapy. MDF has mol.wt. 55-57kD by
CC SDS-PAGE and exhibits an isoelectric point of 6.0-7.0.
XX
SQ Sequence 380 AA;

Query Match 100.0%; Score 83; DB 15; Length 380;
Best Local Similarity 100.0%; Pred. No. 5.7e-05;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIEVTEECTEATATG 17
DB 325 SYIEVTEECTEATATG 341

RESULT 2
AA08254
ID AAY08254 standard; Protein; 380 AA.
XX
AC AAY08254;
XX
DT 14-JUL-1999 (first entry)
XX
DE Human megsin protein.
XX
KW Megsin; mesangial cell; treatment; diagnosis; disease; Iga nephropathy;
KW human; rat; murine.
XX
OS Homo sapiens.
XX
PN WO915652-A1.
XX
PD 01-APR-1999.
XX
PF 22-SEP-1998; 98WO-JP04269.
XX
PR 22-SEP-1997; 97JP-0275302.
XX
PA (KUROO) KUROKAWA K.
PA (MIYA/) MIYATA T.
XX
PI Miyata T;
XX
DR WPI; 1999-276983/23.
DR N-PSDB; AAX56712.
XX
XX Megsin protein expressed specifically in mesangial cells
PS Claim 1; Page 62-64; 100pp; Japanese.
XX
CC This invention describes the isolation of novel megsin nucleic acid and
CC proteins from human, rat and mouse tissue. This protein is expressed
CC specifically in mesangial cells. The products of the invention are
CC useful for the treatment and diagnosis of diseases involving mesangial
CC cells, such as Iga nephropathy.
XX
SQ Sequence 380 AA;

Query Match 100.0%; Score 83; DB 20; Length 380;
Best Local Similarity 100.0%; Pred. No. 5.7e-05;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 SYIEVTEECTEATATG 17
DB 325 SYIEVTEECTEATATG 341

RESULT 3
AAB24142
ID AAB24142 standard; Protein; 380 AA.
XX
AC AAB24142;
XX
DT 30-JAN-2001 (first entry)
XX
DE Human megsin protein sequence SEQ ID NO:2.
XX
KW Megsin; mesangium-predominant gene; serpin regulated; nephropathy;
KW Iga; Immunoglobulin A; detection; renal function; renal disorder;
KW diagnosis; biological sample; blood; urine.
XX
OS Homo sapiens.
XX
PN WO200057189-A1.
XX
PD 28-SEP-2000.
XX
PF 17-MAR-2000; 2000WO-JP01646.
XX
PR 19-MAR-1999; 99JP-0075305.
PR 28-OCT-1999; 99JP-0306623.
XX
PA (KUROO) KUROKAWA K.
PA (FUSO) FUSO PHARM IND LTD.
PA (MIYA/) MIYATA T.
XX
PI Miyata T;
XX
DR WPI; 2000-611642/58.
DR N-PSDB; AAA99294.
XX
XX Evaluating renal function comprises assaying megsin protein in
PT biological sample
XX
PS Example 2; Page 66-69; 93pp; Japanese.
XX
CC The present invention describes a method for evaluating renal function.
CC The method comprises assaying megsin protein in biological sample. Also
CC described are: (1) use of a anti-megsin protein antibody for diagnosing
CC renal function; and (2) a kit for detecting megsin protein comprising:
CC (a) anti-megsin protein antibody attached to solid magnetic particles;
CC (b) direct or indirect fixing for the antibody to the particles; and
CC (c) a magnet. The process is useful for evaluating renal function and
CC diagnosing renal disorders by assaying megsin protein in biological
CC samples (preferably urine or blood). The process is reproducible and
CC gives accurate results. The present sequence represents the human megsin
CC protein, which is given in the exemplification of the present invention.
XX
SQ Sequence 380 AA;

Query Match 100.0%; Score 83; DB 21; Length 380;
Best Local Similarity 100.0%; Pred. No. 5.7e-05;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIEVTEECTEATATG 17
DB 325 SYIEVTEECTEATATG 341

RESULT 4
AAB83075
ID AAB83075 standard; Protein; 380 AA.
AC AAB83075;

```



```

XX 10-JUL-2001 (first entry)
XX Human megsln protein.
DE
XX Human: megsln; mesangial cell proliferative nephritis; nephrotropic;
KW transgenic mouse; glomerular disease; animal model; drug screening.
XX
OS Homo sapiens.
XX
PN WO200124628-A1.
XX
PD 12-APR-2001.
XX
PF 06-OCT-2000; 2000MO-JP06988.
XX
PR 06-OCT-1999; 99JP-0285736.
XX
PA (KURO/) KUROKAWA K.
PA (MIYA/) MIYATA T.
XX
PI Miyata T.
XX
DR WPI; 2001-300136/31.
DR N-PSDB; AAF82438.
XX
PT Mouse model for mesangial cell proliferative nephritis for development
PT and screening of new treatments.
XX
PS Example 4; Page 44-46; 62pp; Japanese.
XX
CC The present sequence is human megsln. The human megsln coding
CC sequence may be introduced into a mouse to produce an animal model of
CC mesangial cell proliferative nephritis. The symptoms include
CC enlargement of the mesangial base region, sedimentation of an immune
CC complex and an increase in mesangial cells. The animal model is useful
CC for analysing the pathology of chronic glomerular diseases and for
CC screening compositions for prevention and treatment of the diseases.
CC Highly uniform models can be made easily and in large numbers using
CC this method.
XX
SQ Sequence 380 AA;
XX
QY 1 SYIEVTEGTEATATG 17
    ||||||||||||
DB 325 SYIEVTEGTEATATG 341

Query Match 100.0%; Score 83; DB 22; Length 380;
Best Local Similarity 100.0%; Pred. No. 5.7e-05;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5
AA57110
ID AAR57110 standard; peptide; 17 AA.
XX
AC AAR57110;
XX
DT 16-AUG-1994 (first entry)
XX
DE Human megakaryocyte differentiation factor peptide 4.
XX
KW Human megakaryocyte differentiation factor; MDF; thrombopoietin;
KW haematopoietic stimulating factor; thrombocytopoiesis; platelet;
KW bone marrow transplantation; cancer chemotherapy.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 1 /note= "not determined"
FT Misc-difference 13 /note= "not determined"
FT Misc-difference 13 /note= "not determined"

```

```

FT Misc-difference 16 /note= "not determined"
FT
XX EP583884-A.
PN
XX 23-FEB-1994.
PD
XX 19-JUL-1993; 93EP-0305654.
PF
XX 17-JUL-1992; 92JP-0212305.
PR 04-MAR-1993; 93JP-0067339.
XX
PA (SUNR ) SUNTORY LTD.
PA (TSUJ/) TSUJIMOTO M.
XX
PI Ishida N, Iwasa F, Kurihara T, Miura K, Nakazato H;
PI Tsujimoto M, Tsuruoka N, Yamaguchi N, Yamachi K;
XX
DR WPI; 1994-058782/08.
XX
PT New megakaryocyte differentiation factor - isolated from human
PT epidermoid carcinoma cells, used to treat conditions involving a
PT decrease in platelets
XX
PS Claim 1; Page 19; 47pp; English.
XX
CC Human MDF (see AAR48379) can be isolated from a culture of human
CC epidermoid carcinoma A431 cells in protein-free medium. The MDF
CC stimulates differentiation of megakaryocytes from myeloid cells
CC in the presence of IL-3. The MDF acts in vivo as a thrombopoietin
CC making it useful for treatment of diseases involving a decrease
CC in platelet number (esp. thrombocytopaenia) such as occurs in bone
CC marrow transplantation and in chemotherapy. MDF has mol.wt.
CC 53-57KD by SDS-PAGE, exhibits an isoelectric point of 6.0-7.0 and
CC contains an amino acid sequence comprising at least one of the
CC sequences AAR57107-R57115.
XX
SQ Sequence 17 AA;
XX
QY 2 YIEVTEGTEATATG 17
    ||||||||||||
DB 2 YIEVTEGTEATATG 17

Query Match 83.1%; Score 69; DB 15; Length 17;
Best Local Similarity 87.5%; Pred. No. 0.00025;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

RESULT 6
AA529329
ID AAY29329 standard; Protein; 217 AA.
XX
AC AAY29329;
XX
DT 29-SEP-1999 (first entry)
XX
DE Human secreted protein clone 1t217_2 protein sequence.
XX
KW Human; secreted protein; nutrition; cytokine; cell proliferation;
KW differentiation; immune stimulating; vaccine; suppression; gene therapy;
KW haematopoiesis regulation; tissue growth; activin; inhibin; cadherin;
KW chemotactic; chemokine; haemostatic; thrombolytic; anti-inflammatory;
KW tumour invasion suppressor; tumour inhibition.
XX
OS Homo sapiens.
XX
PN WO9937674-A1.
XX
PD 29-JUL-1999.
XX
PF 21-JAN-1999; 99WO-US01404.
XX
PR 20-JAN-1999; 99US-0235609.

```

PR 22-JAN-1998; 9805-0072134.  
 XX (GEMV ) GENETICS INST INC.  
 XX  
 PI Agostino MJ, Clark HF, Collins-Racie LA, Fechtel K;  
 PI Jacobs K, Lavallie ER, McCoy JM, Merberg D, Steininger RJ;  
 PI Treacy M, Wong GG;  
 XX  
 DR WPI: 1999-458682/38.  
 DR N-PSDB: AAX90444.  
 XX  
 PT New polynucleotides encoding secreted human proteins derived from,  
 PT e.g. fetal brain potentially used as immunostimulators  
 PS Claim 22: Page 119-120; 139pp; English.  
 XX  
 CC The present sequence represents a human secreted protein. Human secreted  
 CC protein polynucleotides and proteins are predicted to have biological  
 CC activities which would make them suitable for treating, preventing or  
 CC ameliorating medical conditions in humans and animals, although no  
 CC supporting data is given. Suggested activities include nutritional  
 CC activity, cytokine and cell proliferation/differentiation activity,  
 CC immune stimulating (e.g. as vaccines) or suppressing activity,  
 CC haematopoiesis regulating activity, tissue growth activity,  
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory  
 CC activity, cadherin/tumour invasion suppressor activity, and tumour  
 CC inhibition activity. The polynucleotides are also stated to be useful  
 CC for gene therapy.  
 CC  
 SQ Sequence 217 AA:  
 QY  
 Best Local Similarity 83.1%; Score 69; DB 20; Length 217;  
 Matches 13; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 DB 160 SYLDVSECTEAAATG 176  
 1 SYIEVTEGTEATATG 17  
 ||:::|||||  
 160 SYLDVSECTEAAATG 176  
 RESULT 7  
 AAU39055  
 ID AAU39055 standard; Protein: 217 AA.  
 XX  
 AC AAU39055;  
 XX  
 DT 16-JAN-2002 (first entry)  
 XX  
 DE Human secreted protein It217\_2.  
 XX  
 KW Human; secreted protein; antiinflammatory; immunosuppressive;  
 KW neurotropic; neuroprotective; antiarthritic; antimicrobial; vulnery;  
 KW cytostatic; antidiabetic; virucide; antifertility; anticonvulsant;  
 KW vasotropic; antiparkinsonian; immunostimulant; dermatological;  
 KW antirheumatic; antitumor; antitumor; osteopathic; tranquilliser;  
 KW cerebroprotective; cytokine; cell proliferation; cell differentiation;  
 KW immune deficiency; severe combined immunodeficiency; SCID; tumour;  
 KW autoimmune disorder; multiple sclerosis; rheumatoid arthritis;  
 KW graft-versus-host disease; myeloid deficiency; wound healing; ulcer;  
 KW periodontal disease; osteoporosis; osteoarthritis; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; infection; cardiac disease;  
 KW stroke; sepsis; inflammatory bowel disease; contraceptive; immunogen;  
 KW food supplement; vaccine.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175068-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 22-MAR-2001; 2001WO-US09369.  
 XX

PR 30-MAR-2000; 2000US-0539330.  
 PR 04-DEC-2000; 2000US-0729674.  
 XX  
 XX (GEMV ) GENETICS INST INC.  
 XX  
 PI Jacobs K, McCoy JM, Lavallie E, Collins-Racie LA, Evans C;  
 PI Treacy M, Agostino MJ, Steininger RJ, Spaulding V, Wong GG;  
 PI Clark H, Fechtel K, Merberg D;  
 XX  
 DR WPI: 2001-639363/73.  
 DR N-PSDB: AAS99273.  
 XX  
 PT Secreted human proteins, useful as vaccine for treating various  
 PT diseases such as autoimmune disorders (e.g. multiple sclerosis), and  
 PT nervous system disorders (e.g. stroke)  
 PS Disclosure: Page 546-547; 619pp; English.  
 XX  
 CC The invention relates to novel human secreted proteins, the nucleic  
 CC acids encoding them. The protein may exhibit cytokine, cell proliferation  
 CC or cell differentiation activity or may induce production of other  
 CC cytokines in certain cell populations and may exhibit immune stimulating  
 CC or immune suppressing activity, which is useful for the treatment of  
 CC various immune deficiencies and disorders e.g. severe combined  
 CC immunodeficiency (SCID), autoimmune disorders e.g. multiple sclerosis,  
 CC systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary  
 CC inflammation. The proteins are also useful in the treatment of diseases  
 CC and disorders including tissue, skin and organ transplantation and in  
 CC graft-versus-host diseases (GVHD), in the induction of tumour immunity,  
 CC myeloid or lymphoid cell deficiencies, wound healing and tissue repair,  
 CC in the treatment of burns, incisions or ulcers; as well as in treatment  
 CC of periodontal disease, osteoporosis or osteoarthritis, mediated by  
 CC inflammatory processes, diseases of the peripheral nervous system,  
 CC Alzheimer's, Parkinson's disease, Huntington's disease,  
 CC amyotrophic lateral sclerosis, and Shy-Drager syndrome, infections,  
 CC infection of cardiac and central nervous system vessel e.g. stroke,  
 CC sepsis, inflammatory bowel disease, ulcers, bone regeneration. The  
 CC protein, having activin- or inhibin-related activities is useful as a  
 CC contraceptive based on the ability of inhibins to decrease fertility in  
 CC female mammals and decrease spermatogenesis in male mammals. The  
 CC proteins and nucleic acids are also useful as food supplements. The  
 CC present sequence represents a secreted protein of the invention.  
 CC  
 SQ Sequence 217 AA:  
 QY  
 Query Match 83.1%; Score 69; DB 22; Length 217;  
 Best Local Similarity 76.5%; Pred. No. 0.0049;  
 Matches 13; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 DB 160 SYLDVSECTEAAATG 176  
 1 SYIEVTEGTEATATG 17  
 ||:::|||||  
 160 SYLDVSECTEAAATG 176  
 RESULT 8  
 ABB55764  
 ID ABB55764 standard; Protein: 217 AA.  
 XX  
 AC ABB55764;  
 XX  
 DT 14-FEB-2002 (first entry)  
 XX  
 DE Human polypeptide SEQ ID NO 134.  
 XX  
 KW Human; clone bd306-7; clone yb8-1; ATCC number 98599; gene therapy;  
 KW immune disorder; bacterial infection; fungal infection; cancer; tumour;  
 KW autoimmune disorder; systemic lupus erythematosus; wound; ulcer; inhibin;  
 KW osteoporosis; osteoarthritis; nervous system disorder; neuropathy;  
 KW Alzheimer's disease; Parkinson's disease; Huntington's disease; activin;  
 KW haemophilia; cardiac infarction; stroke; sepsis; arthritis; vulnery;  
 KW ischaemia-reperfusion injury; inflammatory bowel disease; chemotactic;  
 KW Crohn's disease; cytostatic; anti-inflammatory; immunomodulator;  
 KW neuroprotective; haemostatic; thrombolytic; anti-inflammatory.

XX Homo sapiens.  
 OS  
 CC  
 XX US2001039335-A1.  
 PN  
 XX  
 PD 08-NOV-2001.  
 PF  
 XX 04-DEC-2000; 2000US-0729674.  
 XX  
 PR 10-APR-1997; 97US-126425P.  
 PR 04-DEC-1997; 97US-067454P.  
 PR 20-DEC-1997; 97US-068379P.  
 PR 02-JAN-1998; 98US-070346P.  
 PR 07-JAN-1998; 98US-070643P.  
 PR 08-JAN-1998; 98US-070755P.  
 PR 13-JAN-1998; 98US-071304P.  
 PR 22-JAN-1998; 98US-072134P.  
 PR 30-JAN-1998; 98US-073095P.  
 PR 18-FEB-1998; 98US-075038P.  
 PR 30-MAR-2000; 2000US-0539330.  
 PR 23-NOV-1998; 98US-0197886.  
 XX  
 PA (JACO/) JACOBS K.  
 PA (MCCO/) MCCOY J M.  
 PA (LAVA/) LAVALLIE E R.  
 PA (COLL/) COLLINS-RACIE L A.  
 PA (EVAN/) EVANS C.  
 PA (MERB/) MERBERG D..  
 PA (TREN/) TREACY M.  
 PA (AGOS/) AGOSTINO M J.  
 PA (STEL/) STEININGER R J.  
 PA (SPAU/) SPAULDING V.  
 PA (WONG/) WONG G G.  
 PA (CLAR/) CLARK H.  
 PA (FECH/) FECHTEL K.  
 XX  
 XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;  
 PI Merberg D, Treacy M, Agostino MJ, Steindinger RJ, Spaulding V;  
 PI Wong GG, Clark H, Fechtel K;  
 XX  
 DR WPI; 2002-040725/05.  
 DR N-PSDB; ABA90942.  
 XX  
 XX New secreted proteins and encoding polynucleotides, useful in gene  
 PT therapies, particularly for preventing or treating autoimmune  
 PT disorders, cancer, graft-versus-host disease, wound, osteoporosis,  
 PT stroke or inflammations  
 XX  
 PS Disclosure: Page 274; 349pp; English.  
 XX  
 CC The invention relates to isolated polynucleotides (ABA90876-ABA90968 and  
 CC ABA90980) and encoded proteins (AB555698-AB555800), especially,  
 CC polynucleotides SEQ ID NO 1 (ABA90876) and SEQ ID NO 19 (ABA90885) and  
 CC proteins SEQ ID NO 2 (AB555698) and SEQ ID NO 20 (AB55707) contained in  
 CC clones bd306-7 and yb8-1 respectively and the clones bd306-7 and yb8-1  
 CC are deposited with the American Type Culture Collection (ATCC) with  
 CC accession number 98599. The polynucleotides and encoded polypeptides have  
 CC cytoskeletal, anti-inflammatory, immunomodulator, vulnerary,  
 CC neuroprotective, activin, inhibin, chemotactic, haemostatic, thrombolytic  
 CC and anti-inflammatory activity and acting as cytokine modulators,  
 CC hematopoiesis regulators, tissue growth modulators and/or cadherin  
 CC suppressors. The polypeptides and polynucleotides are useful in gene  
 CC therapies, particularly for preventing, treating or ameliorating any of  
 CC the following diseases: immune deficiency and disorders; e.g. bacterial  
 CC or fungal infections, autoimmune disorders, cancer, systemic lupus  
 CC erythematosus or graft-versus-host disease; myeloid or lymphoid cell  
 CC deficiencies; wound, burns, incisions and ulcers, osteoporosis or  
 CC osteoarthritis; central and peripheral nervous system diseases and  
 CC neuropathies, e.g. Alzheimer's, Parkinson's disease, Huntington's  
 CC disease, amyotrophic lateral sclerosis or Shy-Drager syndrome;  
 CC haemophilia, cardiac infarction or stroke; Inflammations, shock, sepsis  
 CC or systemic inflammatory response syndrome, ischemia-reperfusion  
 CC injury, endotoxin lethality, arthritis, inflammatory bowel disease or

CC Crohn's disease; or tumours or cancers, pemphigus vulgaris or pemphigus  
 CC foliaceus.  
 CC  
 XX  
 SQ Sequence 217 AA;  
 Query Match 83.1%; Score 69; DB 23; Length 217;  
 Best Local Similarity 76.5%; Pred. No. 0.0049;  
 Matches 13; Conservative 3; Mismatches 1; Indels 0; Caps 0;  
 QY 1 SYIEVTEGTEATATG 17  
 II:::IIIIII IIII  
 Db 160 SYIDVSEGTGTAATG 176  
 RESULT 9  
 AAU16928  
 ID AAU16928 standard; Protein: 340 AA.  
 XX  
 AC AAU16928;  
 XX  
 DT 07-NOV-2001 (first entry)  
 DE Human novel secreted protein, SEQ ID 169.  
 DE  
 XX Human: Immunosuppressive; antiarthritic; antirheumatic;  
 KW cytoskeletal; cardiast; vasotropic; cerebroprotective; nootropic;  
 KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;  
 KW vulnerary; secreted protein; rheumatoid arthritis;  
 KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;  
 KW cerebrovascular disorder; cerebral ischemia; angiogenesis;  
 KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;  
 KW corneal infection; wound healing; epithelial cell proliferation;  
 KW skin ageing; food additive; preservative; antiproliferative.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO20015441-A2.  
 PD  
 XX 02-AUG-2001.  
 XX  
 PE 17-JAN-2001; 2001WO-0501320.  
 XX  
 XX 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198123.  
 PR 19-MAY-2000; 2000US-0205515.  
 PR 07-JUN-2000; 2000US-0209467.  
 PR 28-JUN-2000; 2000US-0214886.  
 PR 30-JUN-2000; 2000US-0215135.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 07-JUL-2000; 2000US-0216880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 11-JUL-2000; 2000US-0217496.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 26-JUL-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.



Matches 13; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 OY 1 SYIEVTEGTEATATG 17  
 |||||:||||| |||||  
 Db 283 SYIEVTEGTEATATG 299

RESULT 10  
 AAY08256  
 ID AAY08256 standard; Protein: 368 AA.

AC AAY08256;

DT 14-JUL-1999 (first entry)

DE Mouse megalin protein.

KW Megalin; mesangial cell; treatment; diagnosis; disease; Iga nephropathy;

KM human; rat; murine.

OS Mus musculus.

FT Key Location/Qualifiers

FT Protein 1..368

PN /note= "partial sequence"

MO9915652-A1.

PD 01-APR-1999.

PF 22-SEP-1998: 98MO-JP04269.

PR 22-SEP-1997: 97JP-0275302.

PA (KURO/) KUROKAWA K.

PA (MIYA/) MIYATA T.

PI Miyata T;

DR WPI: 1999-276983/23.

DR N-PSDB: AAX56714.

PT Megalin protein expressed specifically in mesangial cells

PS Claim 1; Page 76-79; 100pp; Japanese.

CC This invention describes the isolation of novel megalin nucleic acid and

CC proteins from human, rat and mouse tissue. This protein is expressed

CC specifically in mesangial cells. The products of the invention are

CC useful for the treatment and diagnosis of diseases involving mesangial

CC cells, such as Iga nephropathy.

SO Sequence 368 AA;

Query Match 83.1%; Score 69; DB 20; Length 368;

Best Local Similarity 87.5%; Pred. No. 0.009;

Matches 14; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIEVTEGTEATAT 16

|||||:||||| |||||

Db 313 SYIEVSEGTETATAT 328

RESULT 11

ID AAB24151 standard; Protein: 368 AA.

AC AAB24151;

DT 30-JAN-2001 (first entry)

DE Mouse megalin protein sequence SEQ ID NO:21.

XX

XX

XX

KW Megalin; mesangium-predominant gene; serpin regulated; nephropathy;

KM Iga; immunoglobulin A; detection; renal function; renal disorder;

OS Mus musculus.

PN WO200057189-A1.

PD 28-SEP-2000.

PF 17-MAR-2000; 2000WO-JP01646.

PR 19-MAR-1999; 99JP-0075305.

PR 28-OCT-1999; 99JP-0306623.

PA (KURO/) KUROKAWA K.

PA (FUSO ) FUSO PHARM IND LTD.

PA (MIYA/) MIYATA T.

PI Miyata T;

DR WPI: 2000-611642/58.

DR N-PSDB: AAC55239.

PT Evaluating renal function comprises assaying megalin protein in

PT biological sample

PS Disclosure: Page 89-91; 93pp; Japanese.

CC The present invention describes a method for evaluating renal function.

CC The method comprises assaying megalin protein in biological sample. Also

CC described are: (1) use of an anti-megalin protein antibody for diagnosing

CC renal function; and (2) a kit for detecting megalin protein comprising:

CC (a) anti-megalin protein antibody attached to solid magnetic particles;

CC (b) direct or indirect fixing for the antibody to the particles; and

CC (c) a magnet. The process is useful for evaluating renal function and

CC diagnosing renal disorders by assaying megalin protein in biological

CC samples (preferably urine or blood). The process is reproducible and

CC gives accurate results. The present sequence represents the mouse megalin

CC protein, which is given in the exemplification of the present invention.

SO Sequence 368 AA;

Query Match 83.1%; Score 69; DB 21; Length 368;

Best Local Similarity 87.5%; Pred. No. 0.009;

Matches 14; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIEVTEGTEATAT 16

|||||:||||| |||||

Db 313 SYIEVSEGTETATAT 328

RESULT 12

ID AAB83077 standard; Protein: 368 AA.

AC AAB83077;

DT 10-JUL-2001 (first entry)

DE Murine megalin protein.

OS Mus musculus.

KW Mouse; megalin; mesangial cell proliferative nephritis; nephrotropic;

KM Transgenic mouse; glomerular disease; animal model; drug screening.

PA Mus musculus.

PN WO200124628-A1.

PD 12-APR-2001.

DE 06-OCT-2000; 2000WO-JP06988.

XX

XX

XX

```

PR 06-OCT-1999; 99JP-0285736.
XX
XX (KURO/) KUROKAWA K.
PA (MIYA/) MIYATA T.
XX
XX Miyata T;
XX
XX WPI: 2001-300136/31.
DR N-PSDB; AAF82440.
XX
PT Mouse model for mesangial cell proliferative nephritis for development
PT and screening of new treatments
XX
XX Disclosure; Page 52-53; 62pp; Japanese.
XX
XX The present sequence is murine mesgin protein. The human mesgin coding
CC sequence may be introduced into a mouse to produce an animal model of
CC mesangial cell proliferative nephritis. The symptoms include
CC enlargement of the mesangial base region, sedimentation of an immune
CC complex and an increase in mesangial cells. The animal model is useful
CC for analysing the pathology of chronic glomerular diseases and for
CC screening compositions for prevention and treatment of the diseases.
CC Highly uniform models can be made easily and in large numbers using
CC this method.
XX
XX Sequence 368 AA;
SQ
Query Match 83.1%; Score 69; DB 22; Length 368;
Best Local Similarity 87.5%; Pred. No. 0.009;
Matches 14; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIEVTEGTEATTAAT 16
DB 313 SFIEVSECTEATTAAT 328
11:::|||||

RESULT 13
AAV15155
ID AAV15155 standard; Protein; 379 AA.
XX
XX AAV15155;
AC
XX
XX 07-FEB-2000 (first entry)
DT
XX
XX Human prostate serine protease inhibitor (PROSTAPIN).
DE
XX
XX PROSTAPIN; serine protease inhibitor; serpin family; prostate; human;
KM prostate-specific tumour suppressor; apoptosis-inducer; diagnosis;
KW apoptosis-modulator; prognosis; treatment; prostate cancer;
KM metastatic prostate cancer.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH Inhibitory-site 330..347
FT /label= Serine-protease_inhibitory_domain
FT /note= "Also known as reactive-site loop (RSL) and
FT contains a hinge region and a variable region. RSL
FT binds to protease active site during inhibition"
FT 330..338
FT /label= RSL_hinge_region
FT /note= "confers stability to the serpin-protease
FT complex and is highly conserved among the serpins"
FT 344..345
FT /note= "included in the RSL variable region. PROSTAPIN
FT undergoes cleavage at this site during inhibition"
FT
XX
XX WO958560-A2.
XX
XX 18-NOV-1999.
XX
XX 31-MAR-1999; 99WO-US07123.
XX

```

```

PR 31-MAR-1998; 98US-0080167.
PR 15-MAY-1998; 98US-0085720.
XX
XX (UROG-) UROGENESYS INC.
PA (AFAR/) AFAR D E.
PA (HUBE/) HUBERT R S.
PA (LEON/) LEONG K.
PA (RAIT/) RAITANO A B.
PA (SAFE/) SAFFRAN D C.
XX
XX Afar DE, Hubert RS, Leong K, Raitano AB, Saffran DC;
XX
XX WPI: 2000-039068/03.
DR N-PSDB; AA229049.
XX
XX New isolated prostapin gene, used to develop products for, e.g.
PT treating and determining susceptibility to prostate cancer
PT
XX
XX Claim 1; Fig 1; 63pp; English.
XX
XX The present sequence is that of a PROSTAPIN protein which is a member of
CC serine protease inhibitor (serpin) family and expressed in normal
CC prostate and locally confined prostate cancer. The sequence is derived
CC from overlapping sequences of human cDNA clones 11P2A6 and 103. PROSTAPIN
CC contains a reactive-site loop characteristic of serpin family in
CC the carboxy-terminal region. The protein may function as a prostate-
CC specific tumour suppressor, apoptosis-inducer or apoptosis-modulator.
CC The sequence is useful in diagnosing, prognosing, monitoring and
CC treating prostate cancer, particularly advanced stage and metastatic
CC prostate cancer.
XX
XX Sequence 379 AA;
SQ
Query Match 83.1%; Score 69; DB 21; Length 379;
Best Local Similarity 76.5%; Pred. No. 0.0093;
Matches 13; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 SYIEVTEGTEATTAATG 17
DB 322 SYIDVSECTEATAATG 338
11:::|||||

RESULT 14
AAV15156
ID AAV15156 standard; Protein; 617 AA.
XX
XX AAV15156;
AC
XX
XX 07-FEB-2000 (first entry)
DT
XX
XX Human prostate serine protease inhibitor (PROSTAPIN) mutant.
DE
XX
XX PROSTAPIN mutant; serine protease inhibitor; serpin family; prostate;
KM LAPC-9 AD xenograft; androgen dependent LAPC-9 xenograft; cDNA clone 2;
KW diagnosis; prognosis; treatment; prostate cancer; human;
KM metastatic prostate cancer.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH MISC-difference 90
FT /note= "Corresponds to stop codon resulting due to
FT point mutation"
FT
FT MISC-difference 215
FT /note= "Corresponds to stop codon"
FT
FT MISC-difference 225
FT /note= "Corresponds to stop codon"
FT
FT MISC-difference 232
FT /note= "Corresponds to stop codon"
FT
FT MISC-difference 233
FT /note= "Corresponds to stop codon"
FT
FT MISC-difference 289
FT /note= "Corresponds to stop codon"
FT

```

FT MISC-difference 309 /note- "Corresponds to stop codon"  
 FT MISC-difference 318 /note- "Corresponds to stop codon"  
 FT MISC-difference 322 /note- "Corresponds to stop codon"  
 FT MISC-difference 331 /note- "Corresponds to stop codon"  
 FT MISC-difference 332 /note- "Corresponds to stop codon"  
 FT MISC-difference 367 /note- "Corresponds to stop codon"  
 FT MISC-difference 369 /note- "Corresponds to stop codon"  
 FT MISC-difference 374 /note- "Corresponds to stop codon"  
 FT MISC-difference 378 /note- "Corresponds to stop codon"  
 FT MISC-difference 425 /note- "Corresponds to stop codon"  
 FT MISC-difference 430 /note- "Corresponds to stop codon"  
 FT MISC-difference 148 /note- "Corresponds to stop codon"  
 FT MISC-difference 181 /note- "Wild type Met substituted by Thr"  
 FT MISC-difference 188 /note- "Wild type Thr substituted by Ala"  
 FT MISC-difference /note- "Wild type Arg substituted by Trp"  
 FT Region 194.432 /note- "Insertion sequence corresponding to unspliced intron in the coding sequence"  
 FT MISC-difference 518 /note- "Wild type Ile substituted by Thr"  
 FT MISC-difference 528 /note- "Wild type Pro substituted by Ser"  
 FT MISC-difference 528 /note- "Wild type Pro substituted by Ser"  
 XX WO958560-A2.  
 PD 18-NOV-1999.  
 XX 31-MAR-1999; 99WO-US07123.  
 XX 31-MAR-1998; 98US-0080167.  
 PR 15-MAY-1998; 98US-0085720.  
 XX (UROG-) UROGENESYS INC.  
 PA (AFAR/) AFAR D E.  
 PA (HUBE/) HUBERT R S.  
 PA (LEON/) LEONG K.  
 PA (RAIT/) RAITANO A B.  
 PA (SAFF/) SAFFRAN D C.  
 XX Afar DE, Hubert RS, Leong K, Raitano AB, Saffran DC;  
 XX WPI: 2000-039068/03.  
 DR N-PSDB: AA229050.  
 XX New isolated prostate gene, used to develop products for, e.g.  
 PT treating and determining susceptibility to prostate cancer.  
 XX Example 5; Fig 6; 63pp; English.  
 XX The present sequence is a PROSTAPIN (prostate serine protease inhibitor)  
 CC mutant encoded by cDNA clone 2 derived from cDNA library of  
 CC prostate cancer xenograft LAPC-9 AD which was generated from a bone  
 CC tumour biopsy of a patient with hormone-refractory metastatic prostate  
 CC cancer. Clone 2 is a partially unspliced version of wild type PROSTAPIN  
 CC coding sequence with an insertion sequence and several point mutations.  
 CC PROSTAPIN is expressed in normal prostate and locally confined prostate  
 CC cancer, while its expression is lost or attenuated in advanced stage  
 CC cancer. Wild type PROSTAPIN functionality is either altered or completely  
 CC lost in the mutant. PROSTAPIN is useful in diagnosing, prognosing,  
 CC monitoring and treating prostate cancer, particularly advanced stage and

CC metastatic prostate cancer.  
 XX Sequence 617 AA;  
 SO Query Match 83.1%; Score 69; DB 21; Length 617;  
 Best Local Similarity 76.5%; Pred. No. 0.016;  
 Matches 13; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 SYIEVTEGTEATATG 17  
 DB 560 SYLDVSEGTETAAATG 576  
 RESULT 15  
 AAY72661  
 ID AAY72661 standard; Protein: 135 AA.  
 XX AAY72661;  
 AC 31-MAY-2001 (first entry)  
 DT Human headpin (for head and neck serpin) partial protein #2.  
 DE Human; headpin; head and neck serpin; serine protease inhibitor;  
 XX Human; headpin; head and neck serpin; serine protease inhibitor;  
 KW cytosolic; chromosome 18q21.3-22; therapy; cancer.  
 XX Homo sapiens.  
 OS Homo sapiens.  
 XX Key Location/Qualifiers  
 FH MISC-difference 1..2 /note- "Encoded by ATA"  
 FT MISC-difference 1..2 /note- "Encoded by ATA"  
 XX WO200116324-A2.  
 XX 08-MAR-2001.  
 XX 31-AUG-2000; 2000WO-US24214.  
 XX 31-AUG-1999; 99US-0151776.  
 XX (TEXA) UNIV TEXAS SYSTEM.  
 PA Clayman GL, Nakashima T, Spring PM;  
 DR WPI: 2001-169011/17.  
 DR N-PSDB: AAD02741.  
 XX New isolated polynucleotide encoding a headpin (head and neck serpin  
 PT (serine protease inhibitor)) for diagnosing and treating cancers and  
 PT for identifying substances that modulate the headpin activity -  
 XX Claim 21; Page 210; 213pp; English.  
 XX The present sequence is human headpin (for head and neck serpin (serine  
 CC protease inhibitor)) partial protein. Headpin is a differentially  
 CC expressed serine protease inhibitor that belongs to the ov-serpin  
 CC family. It has been grouped into the cluster of serpins located at  
 CC chromosome 18q21.3/18q22. The headpin cDNA and protein are used to  
 CC diagnose and treat cancer, such as of the head and neck, brain, lung,  
 CC liver, spleen, kidney, lymph node, small intestine, pancreas, blood  
 CC cells, colon, stomach, breast, endometrium, prostate, testicle, ovary,  
 CC skin, oesophagus, bone marrow or blood. They are used to identify  
 CC substances that modulate the headpin activity. Antibodies of headpin  
 CC are also used to diagnose cancers using e.g. enzyme linked immunosorbent  
 CC assays (ELISA).  
 SO Sequence 135 AA;  
 QY Query Match 80.7%; Score 67; DB 22; Length 135;  
 Best Local Similarity 76.5%; Pred. No. 0.0058;  
 Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 SYIEVTEGTEATATG 17

Fri Jul 11 15:10:57 2003

us-10-091-442-4.rag

Page 10

Db           |::| | | | | | | | | |  
             78 SFVAVTEEGTEAAATG 94

Search completed: July 11, 2003, 11:54:20  
Job time : 10.7547 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:49:36 ; Search time 3.0395 Seconds  
(without alignments)  
164.563 Million cell updates/sec

Title: US-10-091-442-4

Perfect score: 83

Sequence: 1 SYIEVTEGTEATATG 17

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*

1: /cgn2\_6/ptodata/1/1aa/5A-COMB.pep:\*

2: /cgn2\_6/ptodata/1/1aa/5B-COMB.pep:\*

3: /cgn2\_6/ptodata/1/1aa/6A-COMB.pep:\*

4: /cgn2\_6/ptodata/1/1aa/6B-COMB.pep:\*

5: /cgn2\_6/ptodata/1/1aa/PCTUS-COMB.pep:\*

6: /cgn2\_6/ptodata/1/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	83	100.0	17	US-08-611-977-4	Sequence 4, Appl1
2	83	100.0	380	US-08-472-659-34	Sequence 34, Appl1
3	83	100.0	380	US-08-474-661-34	Sequence 34, Appl1
4	83	100.0	380	US-08-611-977-34	Sequence 34, Appl1
5	69	83.1	17	US-08-472-659-4	Sequence 4, Appl1
6	69	83.1	17	US-08-474-661-4	Sequence 4, Appl1
7	67	80.7	391	US-09-123-912-110	Sequence 110, App
8	67	80.7	391	US-09-643-597-110	Sequence 110, App
9	67	80.7	400	US-09-123-912-112	Sequence 112, App
10	67	80.7	400	US-09-643-597-112	Sequence 112, App
11	62	74.7	376	US-09-200-965-2	Sequence 2, Appl1
12	62	74.7	379	US-08-121-714-4	Sequence 4, Appl1
13	62	74.7	379	US-08-477-108A-4	Sequence 4, Appl1
14	62	74.7	379	US-08-477-112-4	Sequence 4, Appl1
15	62	74.7	379	PCT-US93-08322-4	Sequence 4, Appl1
16	61	73.5	376	US-08-745-985A-12	Sequence 12, Appl1
17	61	73.5	376	US-08-745-985A-24	Sequence 24, Appl1
18	61	73.5	376	US-08-745-985A-30	Sequence 30, Appl1
19	61	73.5	393	US-08-745-995A-20	Sequence 20, Appl1
20	61	73.5	398	US-08-745-995A-26	Sequence 26, Appl1
21	61	73.5	399	US-08-745-995A-8	Sequence 8, Appl1
22	59	71.1	390	US-08-568-147B-2	Sequence 2, Appl1
23	59	71.1	390	US-09-266-910-3	Sequence 3, Appl1
24	59	71.1	390	US-09-266-910-4	Sequence 4, Appl1
25	58	69.9	374	US-08-464-148-2	Sequence 2, Appl1
26	58	69.9	374	US-08-385-500-2	Sequence 2, Appl1
27	58	69.9	374	US-08-846-784-2	Sequence 2, Appl1

28	58	69.9	376	1	US-08-464-148-4	Sequence 4, Appl1
29	58	69.9	376	1	US-08-385-500-4	Sequence 4, Appl1
30	58	69.9	376	1	US-08-846-784-4	Sequence 4, Appl1
31	56	67.5	375	1	US-08-121-714-8	Sequence 8, Appl1
32	56	67.5	375	1	US-08-477-108A-8	Sequence 8, Appl1
33	56	67.5	375	2	US-08-477-112-8	Sequence 8, Appl1
34	56	67.5	375	5	PCT-US93-08322-8	Sequence 8, Appl1
35	56	67.5	407	3	US-08-948-997-3	Sequence 3, Appl1
36	56	67.5	410	4	US-09-348-817A-3	Sequence 3, Appl1
37	55	66.3	376	4	US-08-745-985A-6	Sequence 6, Appl1
38	55	66.3	376	4	US-08-745-985A-36	Sequence 36, Appl1
39	55	66.3	390	4	US-08-745-995A-18	Sequence 18, Appl1
40	55	66.3	397	4	US-08-745-995A-2	Sequence 2, Appl1
41	55	66.3	397	4	US-08-745-995A-32	Sequence 32, Appl1
42	55	66.3	410	3	US-08-948-997-2	Sequence 3, Appl1
43	55	66.3	410	4	US-09-348-817A-2	Sequence 2, Appl1
44	55	66.3	420	4	US-08-745-995A-14	Sequence 14, Appl1
45	54	65.1	382	1	US-07-768-286B-6	Sequence 6, Appl1

#### ALIGNMENTS

RESULT 1  
US-08-611-977-4  
: Sequence 4, Application US/08611977  
: Patent No. 5972886  
GENERAL INFORMATION:  
: APPLICANT: TSUJIMOTO, Masafumi  
: APPLICANT: IWASA, Fuyuki  
: APPLICANT: TSUROOKA, No. 5972886uo  
: APPLICANT: NAKAZATO, Hiroshi  
: APPLICANT: MIURA, Kenju  
: APPLICANT: ISHIDA, No. 5972886uhitro  
: APPLICANT: KURIHARA, Tatsuya  
: APPLICANT: YAMAICHI, Kozo  
: APPLICANT: YAMAGUCHI, No. 5972886om1  
: TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
: NUMBER OF SEQUENCES: 34  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Burns, Doane, Swecker & Mathis  
: STREET: P.O. Box 1404  
: CITY: Alexandria  
: STATE: Virginia  
: COUNTRY: United States  
: ZIP: 22313-1404  
COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/611.977  
: FILING DATE: 06-MAR-1996  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 08/091,028  
: FILING DATE: 14-JUL-1993  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: JP 4-212305  
: FILING DATE: 17-JUL-1992  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: JP 6-067339  
: FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
: NAME: McGowan, Malcolm K.  
: REGISTRATION NUMBER: 39,300  
: REFERENCE/DOCKET NUMBER: 001560-204  
: TELEPHONE: (703) 836-6620  
: TELEFAX: (703) 836-2021  
: INFORMATION FOR SEQ ID NO: 4:  
: SEQUENCE CHARACTERISTICS:

LENGTH: 17 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-611-977-4

Query Match 100.0%; Score 83; DB 2; Length 17;  
Best Local Similarity 100.0%; Pred. No. 1.7e-07;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIEVTEEGTEAATG 17  
DB 1 SYIEVTEEGTEAATG 17

RESULT 2  
US-08-472-659-34  
Sequence 34, Application US/08472659  
Patent No. 5831030

## GENERAL INFORMATION:

APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSUROOKA, No. 5831030uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MURA, Kenju  
APPLICANT: ISHIDA, No. 5831030uhiro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, No. 5831030omi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/472,659  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993  
APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993

ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-248

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-6620  
INFORMATION FOR SEQ ID NO: 34:

SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

MOLECULE TYPE: protein  
US-08-472-659-34

Query Match 100.0%; Score 83; DB 2; Length 380;

Best Local Similarity 100.0%; Pred. No. 6.4e-06;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIEVTEEGTEAATG 17  
DB 325 SYIEVTEEGTEAATG 341

RESULT 3  
US-08-474-661-34  
Sequence 34, Application US/08474661  
Patent No. 5874253

## GENERAL INFORMATION:

APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSUROOKA, No. 5874253uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MURA, Kenju  
APPLICANT: ISHIDA, No. 5874253uhiro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, No. 5874253omi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/474,661  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: REA, TERESA STANER  
REGISTRATION NUMBER: 30,427  
REFERENCE/DOCKET NUMBER: 001560-204

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-6620  
INFORMATION FOR SEQ ID NO: 34:

SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

MOLECULE TYPE: protein  
US-08-474-661-34

Query Match 100.0%; Score 83; DB 2; Length 380;  
Best Local Similarity 100.0%; Pred. No. 6.4e-06;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIEVTEEGTEAATG 17  
DB 325 SYIEVTEEGTEAATG 341

RESULT 4

US-08-611-977-34  
: Sequence 34, Application US/08611977  
: Patent No. 5972886  
: GENERAL INFORMATION:  
: APPLICANT: TSUJIMOTO, Masafumi  
: APPLICANT: IWASA, Fuyuki  
: APPLICANT: TSURODOKA, No. 5972886uo  
: APPLICANT: NAKAZATO, Hiroshi  
: APPLICANT: MIURA, Kenju  
: APPLICANT: ISHIDA, No. 5972886uhiro  
: APPLICANT: KURIHARA, Tatsuya  
: APPLICANT: YAMAGUCHI, Kozo  
: APPLICANT: YAMAGUCHI, No. 5972886cm1  
: TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
: NUMBER OF SEQUENCES: 34  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Burns, Doane, Swecker & Mathis  
: STREET: P.O. Box 1404  
: CITY: Alexandria  
: STATE: Virginia  
: COUNTRY: United States  
: ZIP: 22313-1404  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patentin Release #1.0, Version #1.30  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/611,977  
: FILING DATE: 06-MAR-1996  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 08/091,028  
: FILING DATE: 14-JUL-1993  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: JP 4-212305  
: FILING DATE: 17-JUL-1992  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: JP 6-067339  
: FILING DATE: 04-MAR-1993  
: ATTORNEY/AGENT INFORMATION:  
: NAME: McGowan, Malcolm K.  
: REGISTRATION NUMBER: 39,300  
: REFERENCE/DOCKET NUMBER: 001560-204  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (703) 836-6620  
: TELEFAX: (703) 836-2021  
: INFORMATION FOR SEQ ID NO: 34:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 380 amino acids  
: TYPE: amino acid  
: TOPOLOGY: linear  
: MOLECULE TYPE: protein  
: US-08-611-977-34

Query Match 100.0%; Score 83; DB 2; Length 380;  
Best Local Similarity 100.0%; Pred. No. 6.4e-06;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIEVTEGTEATATG 17  
Db 325 SYIEVTEGTEATATG 341

RESULT 5  
US-08-472-659-4  
: Sequence 4, Application US/08472659  
: Patent No. 5831030  
: GENERAL INFORMATION:  
: APPLICANT: TSUJIMOTO, Masafumi  
: APPLICANT: IWASA, Fuyuki  
: APPLICANT: TSURODOKA, No. 5831030uo  
: APPLICANT: NAKAZATO, Hiroshi

APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. 5831030uhiro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAGUCHI, Kozo  
APPLICANT: YAMAGUCHI, No. 5831030cm1  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/472,659  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-248  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 17 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-472-659-4

Query Match 83.1%; Score 69; DB 2; Length 17;  
Best Local Similarity 87.5%; Pred. No. 3.2e-05;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 YIEVTEGTEATATG 17  
Db 2 YIEVTEGTEATATG 17

RESULT 6  
US-08-474-661-4  
: Sequence 4, Application US/08474661  
: Patent No. 5874253  
: GENERAL INFORMATION:  
: APPLICANT: TSUJIMOTO, Masafumi  
: APPLICANT: IWASA, Fuyuki  
: APPLICANT: TSURODOKA, No. 5874253uo  
: APPLICANT: NAKAZATO, Hiroshi  
: APPLICANT: MIURA, Kenju  
: APPLICANT: ISHIDA, No. 5874253uhiro  
: APPLICANT: KURIHARA, Tatsuya  
: APPLICANT: YAMAGUCHI, Kozo  
: APPLICANT: YAMAGUCHI, No. 5874253cm1  
: TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
: NUMBER OF SEQUENCES: 34

;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Burns, Doane, Swecker & Mathis  
;; STREET: George Mason Bldg., Washington & Prince Sts.  
;; CITY: Alexandria  
;; STATE: Virginia  
;; COUNTRY: United States  
;; ZIP: 22313-1404  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;;  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/474,661  
;; FILING DATE: 07-JUN-1995  
;; CLASSIFICATION: 435  
;; PRIORITY APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/091,028  
;; FILING DATE: 14-JUL-1993  
;; APPLICATION NUMBER: JP 4-212305  
;; FILING DATE: 17-JUL-1992  
;; PRIORITY APPLICATION DATA:  
;; APPLICATION NUMBER: JP 5-067339  
;; FILING DATE: 04-MAR-1993  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: REA, TERESA STANEK  
;; REGISTRATION NUMBER: 30,427  
;; REFERENCE/DOCKET NUMBER: 001560-204  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (703) 836-6620  
;; TELEFAX: (703) 836-6620  
;; INFORMATION FOR SEQ ID NO: 4:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 17 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;;  
US-08-474-661-4

Query Match 83.1%; Score 69; DB 2; Length 17;  
Best Local Similarity 87.5%; Pred. No. 3.2e-05;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 YIEVTEEGTEAATG 17  
DB 2 YIEVTEEGTEAAXAG 17

RESULT 7  
US-09-123-912-110  
; Sequence 110, Application US/09123912A  
; Patent No. 6312695  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Wang, Tonglong  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY OF LUNG CANCER  
; FILE REFERENCE: 210121.455C1  
; CURRENT APPLICATION NUMBER: US/09/123,912A  
; CURRENT FILING DATE: 1998-07-27  
; PRIOR APPLICATION NUMBER: 09/040,802  
; PRIOR FILING DATE: 1998-03-18  
; NUMBER OF SEQ ID NOS: 114  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 110  
; LENGTH: 391  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-123-912-110

Query Match 80.7%; Score 67; DB 4; Length 391;  
Best Local Similarity 76.5%; Pred. No. 0.0027;  
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 SYIEVTEEGTEAATG 17  
DB 334 SEVAVTEEGTEAATG 350

RESULT 8  
US-09-643-597-110  
; Sequence 110, Application US/09643597  
; Patent No. 6426072  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Tonglong  
; APPLICANT: Fan, Liqun  
; APPLICANT: Kalos, Michael D.  
; APPLICANT: Bangur, Chaitanya S.  
; APPLICANT: Hosken, Nancy  
; APPLICANT: Fanger, Gary R.  
; APPLICANT: Li, Samuel X.  
; APPLICANT: Wang, Aljun  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: McNeill, Patricia D.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS OF LUNG CANCER  
; FILE REFERENCE: 210121.455C1  
; CURRENT APPLICATION NUMBER: US/09/643,597  
; CURRENT FILING DATE: 2000-08-21  
; NUMBER OF SEQ ID NOS: 369  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 110  
; LENGTH: 391  
; TYPE: PRT  
; ORGANISM: Homo sapien  
US-09-643-597-110

Query Match 80.7%; Score 67; DB 4; Length 391;  
Best Local Similarity 76.5%; Pred. No. 0.0027;  
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 SYIEVTEEGTEAATG 17  
DB 334 SEVAVTEEGTEAATG 350

RESULT 9  
US-09-123-912-112  
; Sequence 112, Application US/09123912A  
; Patent No. 6312695  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Wang, Tonglong  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY OF LUNG CANCER  
; FILE REFERENCE: 210121.455C1  
; CURRENT APPLICATION NUMBER: US/09/123,912A  
; CURRENT FILING DATE: 1998-07-27  
; PRIOR APPLICATION NUMBER: 09/040,802  
; PRIOR FILING DATE: 1998-03-18  
; NUMBER OF SEQ ID NOS: 114  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 112  
; LENGTH: 400  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-123-912-112

Query Match 80.7%; Score 67; DB 4; Length 400;  
Best Local Similarity 76.5%; Pred. No. 0.0028;  
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```

RESULT 10
US-09-643-597-112
: Sequence 112, Application US/09643597
: Patent No. 6426072
: GENERAL INFORMATION:
: APPLICANT: Wang, Tongtong
: APPLICANT: Fan, Liqun
: APPLICANT: Kalos, Michael D. S.
: APPLICANT: Bangur, Chaltanya S.
: APPLICANT: Hosken, Nancy
: APPLICANT: Fanger, Gary R.
: APPLICANT: Li, Samuel X.
: APPLICANT: Wang, Aijun
: APPLICANT: Skeiky, Yasir A.W.
: APPLICANT: Henderson, Robert A.
: APPLICANT: McNeill, Patricia D.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
: TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
: FILE REFERENCE: 210121.455C11
: CURRENT APPLICATION NUMBER: US/09/643.597
: CURRENT FILING DATE: 2000-08-21
: NUMBER OF SEQ ID NOS: 369
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 112
: LENGTH: 400
: TYPE: PRT
: ORGANISM: Homo sapien
US-09-643-597-112

Query Match      80.7%: Score 67; DB 4; Length 400;
Best Local Similarity 76.5%: Pred. No. 0.0028;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY      1 SYIEVTEEGTEATATG 17
        ||:||||| ||||
Db      343 SFVAVTEEGTEAAATG 359

RESULT 11
US-09-200-965-2
: Sequence 2, Application US/09200965
: Patent No. 6133422
: GENERAL INFORMATION:
: APPLICANT: ROSEN, ET AL.
: TITLE OF INVENTION: Thrombin Inhibitor
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
: ADDRESSEE: CECCHI, STEWART & OLSTEIN
: STREET: 6 BECKER FARM ROAD
: CITY: ROSELAND
: STATE: NEW JERSEY
: COUNTRY: USA
: ZIP: 07068
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5 INCH DISKETTE
: OPERATING SYSTEM: MS-DOS
: SOFTWARE: WORD PERFECT 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/200.965
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/171,817
: FILING DATE: 22-DECEMBER-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: FERRARO, GREGORY D.
: REGISTRATION NUMBER: 36,134
: REFERENCE/DOCKET NUMBER: 325800-47
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 201-994-1700

```

```

TELEFAX: 201-994-1744
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 376 AMINO ACIDS
: TYPE: AMINO ACID.
: STRANDEDNESS:
: TOPOLOGY: LINEAR
: MOLECULE TYPE: PROTEIN
US-09-200-965-2

Query Match      74.7%: Score 62; DB 4; Length 376;
Best Local Similarity 75.0%: Pred. No. 0.017;
Matches 12; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY      1 SYIEVTEEGTEATAT 16
        ||:||||| |||
Db      319 SFVEVNEEGTEAAAT 334

RESULT 12
US-08-121-714-4
: Sequence 4, Application US/08121714
: Patent No. 5470970
: GENERAL INFORMATION:
: APPLICANT: Sager, Ruth
: TITLE OF INVENTION: MASPIN, A NOVEL SERPIN WITH
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson
: STREET: 225 Franklin Street
: CITY: Boston
: STATE: Massachusetts
: COUNTRY: U.S.A.
: ZIP: 02110-2804
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
: OPERATING SYSTEM: IBM PS/2 Model 502 or 55SX
: SOFTWARE: WordPerfect (Version 5.1)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/121,714
: FILING DATE:
: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/938,823
: FILING DATE: 09/01/92
: APPLICATION NUMBER: 07/844,296
: FILING DATE: 02/28/92
: APPLICATION NUMBER: 07/662,216
: FILING DATE: 02/28/91
: ATTORNEY/AGENT INFORMATION:
: NAME: Fraser, Janis K.
: REGISTRATION NUMBER: 34,819
: REFERENCE/DOCKET NUMBER: 00530/072001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 542-5070
: TELEFAX: (617) 542-8906
: TELETYPE: 200154
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 379
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
US-08-121-714-4

Query Match      74.7%: Score 62; DB 1; Length 379;
Best Local Similarity 75.0%: Pred. No. 0.017;
Matches 12; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY      1 SYIEVTEEGTEATAT 16
        ||:||||| |||

```

Db 322 SEVEVEEGTEAAAT 337

## RESULT 13

US-08-477-108A-4  
; Sequence 4, Application US/08477108A

; Patent No. 5801001

; GENERAL INFORMATION:

; APPLICANT: Sager, Ruth

; APPLICANT: Zou, Zhigang

; APPLICANT: Anisowicz, Anthony

; TITLE OF INVENTION: MASPIN, A NOVEL SERPIN WITH

; TITLE OF INVENTION: TUMOR SUPPRESSING ACTIVITY

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish &amp; Richardson P.C.

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: Massachusetts

; COUNTRY: U.S.A.

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

; COMPUTER: IBM PS/2 Model 502 or 555X

; OPERATING SYSTEM: MS-DOS (Version 5.0)

; SOFTWARE: WordPerfect (Version 5.1)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/477,108A

; FILING DATE: June 7, 1995

; CLASSIFICATION: 536

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: 08/121,714

; FILING DATE: 09/01/93

; APPLICATION NUMBER: 07/938,823

; FILING DATE: 09/01/92

; APPLICATION NUMBER: 07/844,296

; FILING DATE: 02/28/92

; APPLICATION NUMBER: 07/662,216

; FILING DATE: 02/28/91

; ATTORNEY/AGENT INFORMATION:

; NAME: Fraser, Janis K.

; REGISTRATION NUMBER: 34,819

; REFERENCE/DOCKET NUMBER: 06570/002002

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617) 542-5070

; TELEFAX: (617) 542-8906

; TELEX: 200154

; INFORMATION FOR SEQ. ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 379

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

US-08-477-108A-4

Query Match 74.7%; Score 62; DB 1; Length 379;

Best Local Similarity 75.0%; Pred. No. 0.017;

Matches 12; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 SYIEVEEGTEAAAT 16

Db 322 SEVEVEEGTEAAAT 337

## RESULT 14

US-08-477-112-4

; Sequence 4, Application US/08477112

; Patent No. 5905023

; GENERAL INFORMATION:

; APPLICANT: Sager, Ruth

; TITLE OF INVENTION: MASPIN, A NOVEL SERPIN WITH

; TITLE OF INVENTION: TUMOR SUPPRESSING ACTIVITY

; NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish &amp; Richardson P.C.

STREET: 225 Franklin Street

CITY: Boston

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

COMPUTER: IBM PS/2 Model 502 or 555X

OPERATING SYSTEM: MS-DOS (Version 5.0)

SOFTWARE: WordPerfect (Version 5.1)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/477,112

FILING DATE:

CLASSIFICATION: 536

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: 08/121,714

FILING DATE: 09/01/93

APPLICATION NUMBER: 07/938,823

FILING DATE: 09/01/92

APPLICATION NUMBER: 07/844,296

FILING DATE: 02/28/92

APPLICATION NUMBER: 07/662,216

FILING DATE: 02/28/91

ATTORNEY/AGENT INFORMATION:

NAME: Fraser, Janis K.

REGISTRATION NUMBER: 34,819

REFERENCE/DOCKET NUMBER: 06570/002003

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 542-5070

TELEFAX: (617) 542-8906

TELEX: 200154

INFORMATION FOR SEQ. ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 379

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

US-08-477-112-4

Query Match 74.7%; Score 62; DB 2; Length 379;

Best Local Similarity 75.0%; Pred. No. 0.017;

Matches 12; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 SYIEVEEGTEAAAT 16

Db 322 SEVEVEEGTEAAAT 337

## RESULT 15

PCT-US93-08322-4

; Sequence 4, Application PC/TUS9308322

; GENERAL INFORMATION:

; APPLICANT: Sager, Ruth

; TITLE OF INVENTION: MASPIN, A NOVEL SERPIN WITH TUMOR SUPPRESSING ACTIVITY

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish &amp; Richardson

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: Massachusetts

; COUNTRY: U.S.A.

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

; COMPUTER: IBM PS/2 Model 502 or 555X

; OPERATING SYSTEM: MS-DOS (Version 5.0)

; SOFTWARE: WordPerfect (Version 5.1)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US93/08322

; FILING DATE:

; CLASSIFICATION:

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 07/938,823  
 FILING DATE: 09/01/92  
 APPLICATION NUMBER: 07/844,296  
 FILING DATE: 02/28/92  
 APPLICATION NUMBER: 07/662,216  
 FILING DATE: 02/28/91  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Fraser, Janis K.  
 REGISTRATION NUMBER: 34,819  
 REFERENCE/DOCKET NUMBER: 00530/072001  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 542-5070  
 TELEFAX: (617) 542-8906  
 TELEX: 200154  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 379  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: linear  
 PCT-US93-08322-4

Query Match 74.7%; Score 62; DB 5; Length 379;  
 Best Local Similarity 75.0%; Pred. No. 0.017;  
 Matches 12; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 SYIEVTEEGTEATAT 16  
 I::I I I I I I I I I  
 Db 322 SPVEVNEEGTEAAAT 337

Search completed: July 11, 2003, 12:02:07  
 Job time : 4.0395 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: July 11, 2003, 11:59:17 ; Search time 4.98337 Seconds  
(Without alignments)  
397.148 Million cell updates/sec

Title: US-10-091-442-4

Perfect score: 83

Sequence: 1 SYIEVTEEGTEATATG 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 445758 seqs, 116419773 residues

Total number of hits satisfying chosen parameters: 445758

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

Published\_Applications\_AA:\*  
1: /cgn2\_6/ptodata/1/pubppa/US07\_NEW\_PUB pep:\*  
2: /cgn2\_6/ptodata/1/pubppa/PC7\_NEW\_PUB pep:\*  
3: /cgn2\_6/ptodata/1/pubppa/US06\_NEW\_PUB pep:\*  
4: /cgn2\_6/ptodata/1/pubppa/US06\_PUBCOMB pep:\*  
5: /cgn2\_6/ptodata/1/pubppa/PC7US\_PUBCOMB pep:\*  
6: /cgn2\_6/ptodata/1/pubppa/US07\_PUBCOMB pep:\*  
7: /cgn2\_6/ptodata/1/pubppa/US08\_NEW\_PUB pep:\*  
8: /cgn2\_6/ptodata/1/pubppa/US08\_PUBCOMB pep:\*  
9: /cgn2\_6/ptodata/1/pubppa/US09\_NEW\_PUB pep:\*  
10: /cgn2\_6/ptodata/1/pubppa/US09\_PUBCOMB pep:\*  
11: /cgn2\_6/ptodata/1/pubppa/US10\_NEW\_PUB pep:\*  
12: /cgn2\_6/ptodata/1/pubppa/US10\_PUBCOMB pep:\*  
13: /cgn2\_6/ptodata/1/pubppa/US60\_NEW\_PUB pep:\*  
14: /cgn2\_6/ptodata/1/pubppa/US60\_PUBCOMB pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	83	100.0	17	9	US-10-091-442-4
2	83	100.0	17	9	US-09-140-719-4
3	83	100.0	380	9	US-10-091-442-34
4	83	100.0	380	10	US-09-140-719-34
5	69	83.1	217	10	US-09-729-674-134
6	69	83.1	340	10	US-09-764-898-169
7	69	83.1	379	9	US-10-012-896-946
8	69	83.1	379	9	US-09-895-814-946
9	69	83.1	392	9	US-10-198-070-62
10	69	83.1	392	9	US-10-198-070-67
11	69	83.1	392	9	US-10-198-070-70
12	69	83.1	617	9	US-10-012-896-947
13	69	83.1	617	9	US-09-895-814-947
14	67	80.7	391	9	US-10-157-031-355
15	67	80.7	391	9	US-09-466-396A-110
16	67	80.7	391	10	US-09-735-705-110
17	67	80.7	391	10	US-09-850-716A-110
18	67	80.7	391	10	US-09-897-778-110
19	67	80.7	400	9	US-09-466-396A-112

20	67	80.7	400	10	US-09-735-705-112	Sequence 112, App
21	67	80.7	400	10	US-09-850-716A-112	Sequence 112, App
22	67	80.7	400	10	US-09-897-778-112	Sequence 112, App
23	62	74.7	437	10	US-09-925-301-1200	Sequence 1200, App
24	62	74.7	439	10	US-09-925-300-1440	Sequence 1440, App
25	59	71.1	377	9	US-10-165-605A-27	Sequence 27, App1
26	59	71.1	377	9	US-09-823-187-31	Sequence 31, App1
27	59	71.1	377	10	US-09-910-430-27	Sequence 27, App1
28	59	71.1	390	7	US-08-731-566-2	Sequence 2, App1
29	59	71.1	390	9	US-08-823-187-27	Sequence 27, App1
30	59	71.1	390	9	US-08-823-187-28	Sequence 28, App1
31	59	71.1	390	9	US-09-823-187-29	Sequence 29, App1
32	59	71.1	390	9	US-09-823-187-30	Sequence 30, App1
33	58	69.9	374	9	US-09-974-298-95	Sequence 95, App1
34	57	68.7	417	9	US-10-174-550-452	Sequence 452, App
35	57	68.7	417	9	US-10-176-758-452	Sequence 452, App
36	57	68.7	417	9	US-10-175-737-452	Sequence 452, App
37	57	68.7	417	9	US-10-173-706-452	Sequence 452, App
38	57	68.7	417	9	US-10-175-738-452	Sequence 452, App
39	57	68.7	417	9	US-10-175-752-452	Sequence 452, App
40	57	68.7	417	9	US-10-176-482-452	Sequence 452, App
41	57	68.7	417	9	US-10-176-757-452	Sequence 452, App
42	57	68.7	417	9	US-10-176-913-452	Sequence 452, App
43	57	68.7	417	9	US-10-180-552-452	Sequence 452, App
44	57	68.7	417	9	US-10-180-557-452	Sequence 452, App
45	57	68.7	417	9	US-10-173-700-452	Sequence 452, App

## ALIGNMENTS

RESULT 1  
US-10-091-442-4  
Sequence 4, Application US/10091442  
Patent No. US20020164711A1  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
IMASA, Fuyuki  
TSURUOKA, NO. US20020164711A1uo  
NAKAZATO, Hiroshi  
MIURA, Kenju  
ISHIDA, No. US20020164711A1uhiro  
KURIHARA, Tatsuya  
YAMAGUCHI, Kozo  
YAMAGUCHI, No. US20020164711A1omi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/091.442  
FILING DATE: 07-Mar-2002  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/140.719  
FILING DATE: 08-AUG-1998  
APPLICATION NUMBER: US 08/474.661  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/091.028  
FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
APPLICATION NUMBER: JP 6-067339

ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 17 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
US-10-091-442-4

Query Match 100.0%; Score 83; DB 9; Length 17;  
Best Local Similarity 100.0%; Pred. No. 4,6e-07;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYEVTEEGTEATATG 17  
Db 1 SYEVTEEGTEATATG 17

RESULT 2  
US-09-140-719-4  
Sequence 4, Application US/09140719  
Patent No. US20010026931A1  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IMASA, Fuyuki  
APPLICANT: TSUROUOKA, No. US20010026931A1uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. US20010026931A1uhiro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, No. US20010026931A1om1  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.

REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 17 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-140-719-4

Query Match 100.0%; Score 83; DB 10; Length 17;  
Best Local Similarity 100.0%; Pred. No. 4,6e-07;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYEVTEEGTEATATG 17  
Db 1 SYEVTEEGTEATATG 17

RESULT 3  
US-10-091-442-34  
Sequence 34, Application US/10091442  
Patent No. US20020164711A1  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IMASA, Fuyuki  
APPLICANT: TSUROUOKA, No. US20020164711A1uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. US20020164711A1uhiro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, No. US20020164711A1om1  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/091,442  
FILING DATE: 07-MAR-2002  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021

;; INFORMATION FOR SEQ ID NO: 34:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 380 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; SEQUENCE DESCRIPTION: SEQ ID NO: 34:  
US-10-091-442-34

Query Match 100.0%; Score 83; DB 9; Length 380;  
Best Local Similarity 100.0%; Pred. No. 1.4e-05;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIEVTEEGTEAATG 17  
DB 325 SYIEVTEEGTEAATG 341

RESULT 4  
US-09-140-719-34  
; Sequence 34, Application US/09140719  
; Patent No. US20010026931A1  
; GENERAL INFORMATION:  
; APPLICANT: TSUJIMOTO, Masafumi  
; APPLICANT: IMASA, Fuyuki  
; APPLICANT: TSUROUOKA, No. US20010026931A1uo  
; APPLICANT: MAKAZATO, Hiroshi  
; APPLICANT: MIURA, Kenju  
; APPLICANT: ISHIDA, No. US20010026931A1uhro  
; APPLICANT: KURIHARA, Tatsuya  
; APPLICANT: YAMAICHI, Kozo  
; APPLICANT: YAMAGUCHI, No. US20010026931A1oml  
; TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
; NUMBER OF SEQUENCES: 34  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: P.O. Box 1404  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/140,719  
; FILING DATE: 08-AUG-1998  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/474,661  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/091,028  
; FILING DATE: 14-JUL-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 4-212305  
; FILING DATE: 17-JUL-1992  
; APPLICATION DATA:  
; APPLICATION NUMBER: JP 6-067339  
; FILING DATE: 04-MAR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McGowan, Malcolm K.  
; REGISTRATION NUMBER: 39,300  
; REFERENCE/DOCKET NUMBER: 001560-247  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 34:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 380 amino acids  
; TYPE: amino acid

;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-09-140-719-34

Query Match 100.0%; Score 83; DB 10; Length 380;  
Best Local Similarity 100.0%; Pred. No. 1.4e-05;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIEVTEEGTEAATG 17  
DB 325 SYIEVTEEGTEAATG 341

RESULT 5  
US-09-729-674-134  
; Sequence 134, Application US/09729674  
; Patent No. US20010039335A1  
; GENERAL INFORMATION:  
; APPLICANT: Jacobs, Kenneth  
; APPLICANT: McCoy, John M.  
; APPLICANT: Lavallee, Edward R.  
; APPLICANT: Collins-Racie, Lisa A.  
; APPLICANT: Evans, Cheryl  
; APPLICANT: Merberg, David  
; APPLICANT: Treacy, Maurice  
; APPLICANT: Agostino, Michael J.  
; APPLICANT: Steininger II, Robert J.  
; APPLICANT: Spaulding, Vikki  
; APPLICANT: Wong, Gordon G.  
; APPLICANT: Clark, Hilary  
; APPLICANT: Fectchel, Kim  
; APPLICANT: Genetics-Institute, Inc.  
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM  
; FILE REFERENCE: 6035-64X  
; CURRENT APPLICATION NUMBER: US/09/729,674  
; CURRENT FILING DATE: 2000-12-04  
; PRIOR APPLICATION NUMBER: 09/539,330  
; PRIOR FILING DATE: 2000-03-30  
; NUMBER OF SEQ ID NOS: 283  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 134  
; LENGTH: 217  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-729-674-134

Query Match 83.1%; Score 69; DB 10; Length 217;  
Best Local Similarity 76.5%; Pred. No. 0.0014;  
Matches 13; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 SYIEVTEEGTEAATG 17  
DB 160 SYLDVSEEGTEAAATG 176

RESULT 6  
US-09-764-898-169  
; Sequence 169, Application US/09764898  
; Patent No. US20020090673A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, proteins, and Antibodies  
; FILE REFERENCE: P0201  
; CURRENT APPLICATION NUMBER: US/09/764,898  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 311  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 169  
; LENGTH: 340  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-764-898-169

Query Match 83.1%; Score 69; DB 10; Length 340;  
 Best Local Similarity 76.5%; Pred. No. 0.0023;  
 Matches 13; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 SYIEVTEECTEATATG 17  
 ||:::||||| ||||  
 Db 283 SYLDVSECTEAAATG 299

## RESULT 7

US-10-012-896-946

; Sequence 946, Application US/10012896  
 ; Publication No. US20020183251A1  
 ; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun  
 ; APPLICANT: Dillon, Davin C.  
 ; APPLICANT: Mitcham, Jennifer L.  
 ; APPLICANT: Harlocker, Susan L.  
 ; APPLICANT: Jiang, Yugu  
 ; APPLICANT: Kalos, Michael D.  
 ; APPLICANT: Retter, Marc W.  
 ; APPLICANT: Stolk, John A.  
 ; APPLICANT: Day, Craig H.  
 ; APPLICANT: Vedvick, Thomas S.  
 ; APPLICANT: Carter, Darrick  
 ; APPLICANT: Li, Samuel X.  
 ; APPLICANT: Wang, Aljun  
 ; APPLICANT: Skeiky, Yasir A.W.  
 ; APPLICANT: Hepler, William T.  
 ; APPLICANT: Henderson, Robert A.  
 ; APPLICANT: Hural, John  
 ; APPLICANT: McNeill, Patricia D.  
 ; APPLICANT: Houghton, Raymond L.  
 ; APPLICANT: Vinals de Bassols, Carlota  
 ; APPLICANT: Foy, Teresa  
 ; APPLICANT: Fanger, Gary R.  
 ; APPLICANT: Mantanabe, Yoshihiro  
 ; APPLICANT: Meagher, Madeleine Joy  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
 ; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER  
 ; FILE REFERENCE: 210121.427C27  
 ; CURRENT APPLICATION NUMBER: US/10/012,896  
 ; CURRENT FILING DATE: 2001-12-10  
 ; NUMBER OF SEQ ID NOS: 1011  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 946  
 ; LENGTH: 379  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-10-012-896-946

Query Match 83.1%; Score 69; DB 9; Length 379;  
 Best Local Similarity 76.5%; Pred. No. 0.0026;  
 Matches 13; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 SYIEVTEECTEATATG 17  
 ||:::||||| ||||  
 Db 322 SYLDVSECTEAAATG 338

## RESULT 8

US-09-895-814-946

; Sequence 946, Application US/09895814  
 ; Publication No. US20020193296A1  
 ; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun  
 ; APPLICANT: Dillon, Davin C.  
 ; APPLICANT: Mitcham, Jennifer L.  
 ; APPLICANT: Harlocker, Susan L.  
 ; APPLICANT: Jiang, Yugu  
 ; APPLICANT: Kalos, Michael D.  
 ; APPLICANT: Retter, Marc W.

; APPLICANT: Stolk, John A.  
 ; APPLICANT: Day, Craig H.  
 ; APPLICANT: Vedvick, Thomas S.  
 ; APPLICANT: Carter, Darrick  
 ; APPLICANT: Li, Samuel X.  
 ; APPLICANT: Wang, Aljun  
 ; APPLICANT: Skeiky, Yasir A.W.  
 ; APPLICANT: Hepler, William T.  
 ; APPLICANT: Henderson, Robert A.  
 ; APPLICANT: Hural, John  
 ; APPLICANT: McNeill, Patricia D.  
 ; APPLICANT: Houghton, Raymond L.  
 ; APPLICANT: Vinals de Bassols, Carlota  
 ; APPLICANT: Foy, Teresa  
 ; APPLICANT: Fanger, Gary R.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
 ; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER  
 ; FILE REFERENCE: 210121.427C26  
 ; CURRENT APPLICATION NUMBER: US/09/895,814  
 ; CURRENT FILING DATE: 2001-06-29  
 ; NUMBER OF SEQ ID NOS: 990  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 946  
 ; LENGTH: 379  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-09-895-814-946

Query Match 83.1%; Score 69; DB 9; Length 379;  
 Best Local Similarity 76.5%; Pred. No. 0.0026;  
 Matches 13; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 SYIEVTEECTEATATG 17  
 ||:::||||| ||||  
 Db 322 SYLDVSECTEAAATG 338

## RESULT 9

US-10-198-070-62

; Sequence 62, Application US/10198070  
 ; Publication No. US20030109437A1  
 ; GENERAL INFORMATION:

; APPLICANT: AVERBACK, PAUL  
 ; APPLICANT: GEMMELL, JACK  
 ; TITLE OF INVENTION: PEPTIDES EFFECTIVE IN THE TREATMENT OF TUMORS AND OTHER  
 ; TITLE OF INVENTION: CONDITIONS REQUIRING THE REMOVAL OR DESTRUCTION OF  
 ; TITLE OF INVENTION: CELLS  
 ; FILE REFERENCE: 59003.000008  
 ; CURRENT APPLICATION NUMBER: US/10/198,070  
 ; CURRENT FILING DATE: 2002-07-19  
 ; PRIOR APPLICATION NUMBER: 60/306,161  
 ; PRIOR FILING DATE: 2001-07-19  
 ; PRIOR APPLICATION NUMBER: 60/306,150  
 ; PRIOR FILING DATE: 2001-07-19  
 ; PRIOR APPLICATION NUMBER: 60/331,477  
 ; PRIOR FILING DATE: 2001-11-16  
 ; NUMBER OF SEQ ID NOS: 125  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 62  
 ; LENGTH: 392  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-10-198-070-62

Query Match 83.1%; Score 69; DB 9; Length 392;  
 Best Local Similarity 76.5%; Pred. No. 0.0027;  
 Matches 13; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 SYIEVTEECTEATATG 17  
 ||:::||||| ||||  
 Db 335 SYLDVSECTEAAATG 351

RESULT 10  
US-10-198-070-67  
: Sequence 67, Application US/10198070  
: Publication No. US20030109437A1  
: GENERAL INFORMATION:  
: APPLICANT: AVERBACK, PAUL  
: APPLICANT: GEMMELL, JACK  
: TITLE OF INVENTION: PEPTIDES EFFECTIVE IN THE TREATMENT OF TUMORS AND OTHER  
: TITLE OF INVENTION: CONDITIONS REQUIRING THE REMOVAL OR DESTRUCTION OF  
: TITLE OF INVENTION: CELLS  
: FILE REFERENCE: 59003.000008  
: CURRENT APPLICATION NUMBER: US/10/198.070  
: CURRENT FILING DATE: 2002-07-19  
: PRIOR APPLICATION NUMBER: 60/306,161  
: PRIOR FILING DATE: 2001-07-19  
: PRIOR APPLICATION NUMBER: 60/306,150  
: PRIOR FILING DATE: 2001-07-19  
: PRIOR APPLICATION NUMBER: 60/331,477  
: PRIOR FILING DATE: 2001-11-16  
: NUMBER OF SEQ ID NOS: 125  
: SOFTWARE: PatentIn Ver. 2.1  
: SEQ ID NO 67  
: LENGTH: 392  
: TYPE: PRT  
: ORGANISM: Homo sapiens  
US-10-198-070-67

Query Match 83.1%, Score 69; DB 9; Length 392;  
Best Local Similarity 76.5%; Pred. No. 0.0027;  
Matches 13; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Oy 1 SYIEVTEGTEATAATG 17  
||:::|||||  
Db 335 SYLDVSEGTAAATG 351

RESULT 11  
US-10-198-070-70  
: Sequence 70, Application US/10198070  
: Publication No. US20030109437A1  
: GENERAL INFORMATION:  
: APPLICANT: AVERBACK, PAUL  
: APPLICANT: GEMMELL, JACK  
: TITLE OF INVENTION: PEPTIDES EFFECTIVE IN THE TREATMENT OF TUMORS AND OTHER  
: TITLE OF INVENTION: CONDITIONS REQUIRING THE REMOVAL OR DESTRUCTION OF  
: TITLE OF INVENTION: CELLS  
: FILE REFERENCE: 59003.000008  
: CURRENT APPLICATION NUMBER: US/10/198.070  
: CURRENT FILING DATE: 2002-07-19  
: PRIOR APPLICATION NUMBER: 60/306,161  
: PRIOR FILING DATE: 2001-07-19  
: PRIOR APPLICATION NUMBER: 60/306,150  
: PRIOR FILING DATE: 2001-07-19  
: PRIOR APPLICATION NUMBER: 60/331,477  
: PRIOR FILING DATE: 2001-11-16  
: NUMBER OF SEQ ID NOS: 125  
: SOFTWARE: PatentIn Ver. 2.1  
: SEQ ID NO 70  
: LENGTH: 392  
: TYPE: PRT  
: ORGANISM: Homo sapiens  
US-10-198-070-70

Query Match 83.1%, Score 69; DB 9; Length 392;  
Best Local Similarity 76.5%; Pred. No. 0.0027;  
Matches 13; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Oy 1 SYIEVTEGTEATAATG 17  
||:::|||||  
Db 335 SYLDVSEGTAAATG 351

RESULT 12

US-10-012-896-947  
: Sequence 947, Application US/10012896  
: Publication No. US20020183251A1  
: GENERAL INFORMATION:  
: APPLICANT: Xu, Jiangchun  
: APPLICANT: Dillon, Davin C.  
: APPLICANT: Mitcham, Jennifer L.  
: APPLICANT: Harlocker, Susan L.  
: APPLICANT: Jiang, Yugu  
: APPLICANT: Kalos, Michael D.  
: APPLICANT: Retter, Marc W.  
: APPLICANT: Stolck, John A.  
: APPLICANT: Day, Craig H.  
: APPLICANT: Vedvick, Thomas S.  
: APPLICANT: Carter, Darlick  
: APPLICANT: Li, Samuel X.  
: APPLICANT: Wang, Aljun  
: APPLICANT: Skeiky, Yasir A.W.  
: APPLICANT: Hepler, William T.  
: APPLICANT: Henderson, Robert A.  
: APPLICANT: Hural, John  
: APPLICANT: McNeill, Patricia D.  
: APPLICANT: Houghton, Raymond L.  
: APPLICANT: Vinals de Baassols, Carlota  
: APPLICANT: Foy, Teresa  
: APPLICANT: Fanger, Gary R.  
: APPLICANT: Mantanabe, Yoshihiro  
: APPLICANT: Meagher, Madeline Joy  
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
: TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER  
: FILE REFERENCE: 210121.427C27  
: CURRENT APPLICATION NUMBER: US/10/012.896  
: CURRENT FILING DATE: 2001-12-10  
: NUMBER OF SEQ ID NOS: 1011  
: SOFTWARE: FastSeq for Windows Version 3.0  
: SEQ ID NO 947  
: LENGTH: 617  
: TYPE: PRT  
: ORGANISM: Homo sapiens  
: FEATURE:  
: NAME/KEY: VARIANT  
: LOCATION: 90, 215, 225, 232, 233, 289, 309, 318, 322, 331, 332, 367,  
: LOCATION: 369, 374, 378, 425, 430  
: OTHER INFORMATION: Xaa - Any Amino Acid  
US-10-012-896-947

Query Match 83.1%, Score 69; DB 9; Length 617;  
Best Local Similarity 76.5%; Pred. No. 0.0045;  
Matches 13; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Oy 1 SYIEVTEGTEATAATG 17  
||:::|||||  
Db 560 SYLDVSEGTAAATG 576

RESULT 13  
US-09-895-814-947  
: Sequence 947, Application US/09895814  
: Publication No. US20020193296A1  
: GENERAL INFORMATION:  
: APPLICANT: Xu, Jiangchun  
: APPLICANT: Dillon, Davin C.  
: APPLICANT: Mitcham, Jennifer L.  
: APPLICANT: Harlocker, Susan L.  
: APPLICANT: Jiang, Yugu  
: APPLICANT: Kalos, Michael D.  
: APPLICANT: Retter, Marc W.  
: APPLICANT: Stolck, John A.  
: APPLICANT: Day, Craig H.  
: APPLICANT: Vedvick, Thomas S.  
: APPLICANT: Carter, Darlick  
: APPLICANT: Li, Samuel X.  
: APPLICANT: Wang, Aljun

```
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C26
; CURRENT APPLICATION NUMBER: US/09/895,814
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 990
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 947
; LENGTH: 617
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: variant
; LOCATION: (1)..(617)
; OTHER INFORMATION: Xaa - Any amino acid
US-09-895-814-947
```

```
Query Match      83.1%; Score 69; DB 9; Length 617;
Best Local Similarity 76.5%; Pred. No. 0.0045;
Matches 13; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 SYIEVTEGTEATATG 17
        11::1111111111111111
DB      560 SYLDVSECTEAAATG 576
```

```
RESULT 14
US-10-157-031-355
; Sequence 355, Application US/10157031
; Publication No. US20030108890A1
; GENERAL INFORMATION:
; APPLICANT: Baranova, A. V.
; APPLICANT: Yankovsky, N. K.
; APPLICANT: Kozlov, A. P.
; APPLICANT: Lobashev, A. V.
; APPLICANT: Krivoskaya, L. L.
; TITLE OF INVENTION: In silico screening for phenotype-associated expressed sequences
; FILE REFERENCE: 2760-103
; CURRENT APPLICATION NUMBER: US/10/157,031
; CURRENT FILING DATE: 2002-05-30
; NUMBER OF SEQ ID NOS: 415
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 355
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-157-031-355
```

```
Query Match      80.7%; Score 67; DB 9; Length 391;
Best Local Similarity 76.5%; Pred. No. 0.0057;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      1 SYIEVTEGTEATATG 17
        1::1111111111111111
DB      334 SEVAVTEGTEAAATG 350
```

```
RESULT 15
US-09-466-396A-110
; Sequence 110, Application US/09466396A
; Publication No. US20030119763A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY AND
```

```
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C4
; CURRENT APPLICATION NUMBER: US/09/466,396A
; CURRENT FILING DATE: 1999-12-17
; NUMBER OF SEQ ID NOS: 224
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 110
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-466-396A-110
```

```
Query Match      80.7%; Score 67; DB 9; Length 391;
Best Local Similarity 76.5%; Pred. No. 0.0057;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      1 SYIEVTEGTEATATG 17
        1::1111111111111111
DB      334 SEVAVTEGTEAAATG 350
```

```
Search completed: July 11, 2003, 12:37:33
Job time : 5.98337 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:45:06 ; Search time 3.42827 Seconds  
(without alignments)  
476.708 Million cell updates/sec

Title: US-10-091-442-4

Perfect score: 83

Sequence: 1 SYIEVTEGTEAATG 17

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR\_73:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	67	80.7	391	2 JC7118	headpin serine pro
2	67	74.7	376	1 A48681	placental thrombin
3	62	74.7	376	2 S38962	serpin - pig
4	60	72.3	379	2 S27383	elastase inhibitor
5	60	72.3	388	1 DYCH	ovalbumin-related
6	59	72.3	399	1 DKBH2	protein 24 - barley
7	59	71.1	378	2 A57488	proteinase inhibitor
8	59	71.1	390	2 I38201	squamous cell carc
9	59	71.1	390	2 I38202	leupin precursor -
10	58	69.9	374	2 A59273	proteinase inhibitor
11	58	69.9	376	2 B59273	proteinase inhibitor
12	58	69.9	397	2 T06183	serpin - barley
13	57	68.7	374	2 A41903	hypothetical prote
14	57	68.7	398	2 S39819	serpin - barley
15	56	67.5	213	2 A26423	serine proteinase
16	56	67.5	410	2 A42421	leukocyte elastase
17	56	67.5	410	2 S70647	neuroserpin precur
18	56	67.5	418	1 S31507	serine proteinase
19	55	66.3	403	1 S08102	serine proteinase
20	54	65.1	407	2 D84518	probable serpin (i
21	54	65.1	415	2 A32853	plasmaogen activa
22	53	63.9	397	2 I39184	hypothetical prote
23	53	63.9	398	2 S65782	homopain - human
24	53	63.9	398	2 S65782	serpin - wheat
25	53	63.9	399	2 T06488	serpin W2S2 - whea
26	53	63.9	429	2 S49595	probable serine pr
27	52	62.7	93	2 S62390	alpha 1 antichymot
28	52	62.7	222	1 WNVZ84	antithrombin-III h
29	52	62.7	252	2 S66289	alpha 1 antichymot

30	52	62.7	341	1 WNVZHI	antithrombin-III h
31	52	62.7	344	2 S46871	B1R protein - var
32	52	62.7	344	2 E72173	D2R protein - vari
33	52	62.7	344	2 T28611	hypothetical prote
34	52	62.7	345	1 WNVZWI	antithrombin-III h
35	52	62.7	391	2 H96517	protein T2E6.22 (1
36	51	61.4	64	2 S15635	serpin Sp12 (clone
37	51	61.4	64	2 S15633	serpin Sp12 (clone
38	51	61.4	66	2 S15629	serpin Sp12 (clone
39	51	61.4	216	2 B26423	serine proteinase
40	51	61.4	344	1 S24676	serpin - Ectromeli
41	51	61.4	353	1 WNVZS1	antithrombin-III h
42	51	61.4	353	1 WNVZS1	antithrombin-III h
43	51	61.4	366	2 T16119	hypothetical prote
44	51	61.4	369	2 JH0493	alpha-1 antichymot
45	51	61.4	397	2 I48717	proteinase inhibitor

#### ALIGNMENTS

```

RESULT 1
JC7118
headpin serine proteinase inhibitor - human
C:Species: Homo sapiens (man)
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 26-May-2000
C:Accession: JC7118
R:Spring, P.; Nakashima, T.; Frederick, M.; Henderson, Y.; Clayman, G.
Biochem. Biophys. Res. Commun. 264, 299-304, 1999
A>Title: Identification and CDNA cloning of headpin, a novel differentially expressed
A:Reference number: JC7118; MUID:99458661; PMID:10527881
A:Accession: JC7118
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-391 <SPR>
A:Cross-references: GB:A5169949; NID:95911368; PIDN:A055765.1; PID:95911369
C:Genetics:
A:Map position: 18q21.3-18q22
C:Superfamily: antithrombin III
C:Keywords: serine proteinase inhibitor
F:355/Inhibitory site: Val (unidentified proteinase) #status predicted

Query Match
Best Local Similarity 80.7%; Score 67; DB 2; Length 391;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Oy 1 SYIEVTEGTEAATG 17
   ||| ||||| ||||
Db 334 SFVAWTEGTEAATG 350

RESULT 2
A48681
placental thrombin inhibitor - human
N:Alternate names: cytoplasmic antiproteinase; intracellular serine proteinase inhibi
C:Species: Homo sapiens (man)
C:Date: 07-Apr-1994 #sequence_revision 07-Jul-1995 #text_change 18-Jun-1999
C:Accession: A48681; A54352; A46672; C46672; S35750
R:Coughlin, P.; Sun, J.; Cerruti, L.; Salem, H.H.; Bird, P.
Proc. Natl. Acad. Sci. U.S.A. 90, 9417-9421, 1993
A>Title: Cloning and molecular characterization of a human intracellular serine prote
A:Reference number: A48681; MUID:94022286; PMID:8415716
A:Accession: A48681
A:Molecule type: mRNA
A:Residues: 1-376 <COO>
A:Cross-references: GB:222658; NID:9297411; PIDN:CAA80373.1; PID:9297412
A:Experimental source: placenta
A>Note: authors translated the codon CAA for residue 198 as Gly
R:Morgenstern, K.A.; Sprecher, C.; Holth, L.; Foster, D.; Grant, F.J.; Ching, A.; Kis
Biochemistry 33, 3432-3441, 1994
A>Title: Complementary DNA cloning and kinetic characterization of a novel intracellu
A:Reference number: A54352; MUID:94183847; PMID:8136380
A:Accession: A54352

```

A:Molecule type: mRNA  
 A:Residues: 1-174, 'E', 176-361, 'S', 363-376 <MOR>  
 A:Cross-references: GB:S69272; NID:9546087; PIDN:AA30320.1; PID:9546088  
 A:Experimental source: Placenta  
 A>Note: sequence extracted from NCBI backbone (NCB1N:145231, NCB1P:145232)  
 R.Coughlin, P.B.; Tetaz, T.; Salem, H.H.  
 J. Biol. Chem. 268, 9341-9347, 1993  
 A:Title: Identification and purification of a novel serine proteinase inhibitor.  
 A:Reference number: A46672; MUID:93252826; PMID:8486644  
 A:Accession: A46672  
 A:Molecule type: protein  
 A:Residues: 47-60;63-81;91-98 <CO2>  
 A:Experimental source: Placenta, leukemic cell line K562  
 A>Note: sequence modified after extraction from NCBI backbone  
 C:Genetics:  
 A:Gene: GDB:P16  
 A:Cross-references: GDB:252025; OMIM:173321  
 A:Map position: 6p25-6p24.3  
 C:Superfamily: antithrombin III  
 C:Keywords: blocked amino end; cytosol; serine proteinase inhibitor  
 F:341/inhibitory site: Arg (thrombin) #status predicted

Query Match 74.7%; Score 62; DB 1; Length 376;  
 Best Local Similarity 75.0%; Pred. No. 0.015;  
 Matches 12; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 SYIEVTEGTEATAT 16  
 I:::| | | | | | | |  
 Db 319 SFVEVNEGTEAAAT 334

RESULT 3  
 Serpin - pig  
 C:Species: Sus scrofa domestica (domestic pig)  
 C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 02-Jul-1998  
 C:Accession: S38962  
 R:Reschauer, W.F.; Mentele, R.; Sommerhoff, C.P.  
 Eur. J. Biochem. 217, 519-526, 1993  
 A:Title: Primary structure of a porcine leukocyte serpin.  
 A:Reference number: S38962; MUID:94039085; PMID:7901009  
 A:Accession: S38962  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 1-378 <RES>  
 A>Note: the sequence from Fig. 6 is inconsistent with that from Fig. 5 in having 256-Asn  
 C:Superfamily: antithrombin III

Query Match 74.7%; Score 62; DB 2; Length 378;  
 Best Local Similarity 75.0%; Pred. No. 0.015;  
 Matches 12; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 SYIEVTEGTEATAT 16  
 I:::| | | | | | | |  
 Db 321 SFVEVNEGTEAAAT 336

RESULT 4  
 elastase inhibitor - human  
 C:Species: Homo sapiens (man)  
 C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 05-Sep-1997  
 C:Accession: S27383; S65750  
 R:Remold-O'Donnell, E.; Chin, J.; Alberts, M.  
 Proc. Natl. Acad. Sci. U.S.A. 89, 5635-5639, 1992  
 A:Title: Sequence and molecular characterization of human monocyte/neutrophil elastase  
 A:Reference number: S27383; MUID:92302296; PMID:1376927  
 A:Accession: S27383  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-379 <REM>  
 A:Cross-references: EMBL:M93056  
 R:Packard, B.Z.; Lee, S.S.; Remold-O'Donnell, E.; Komoriya, A.

Biochim. Biophys. Acta 1269, 41-50, 1995  
 A:Title: A serpin from human tumor cells with direct lymphoid immunomodulatory activity  
 A:Reference number: S65750; MUID:96049524; PMID:7578269  
 A:Accession: S65750  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 57-69;97-110;111-129;204-213;216-244;255-271, 'X', 273-274;291-301 <PAC>  
 C:Genetics:  
 A:Gene: GDB:ELANH2; EI; P12  
 A:Cross-references: GDB:132914; OMIM:130135  
 A:Map position: 6p25-6p24.3  
 C:Superfamily: antithrombin III  
 C:Keywords: serine proteinase inhibitor

Query Match 74.7%; Score 62; DB 2; Length 379;  
 Best Local Similarity 75.0%; Pred. No. 0.015;  
 Matches 12; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 SYIEVTEGTEATAT 16  
 I:::| | | | | | | |  
 Db 322 SFVEVNEGTEAAAT 337

RESULT 5  
 DYCH  
 ovalbumin-related y protein - chicken  
 C:Species: Gallus gallus (chicken)  
 C:Date: 17-Dec-1982 #sequence\_revision 17-Dec-1982 #text\_change 18-Jun-1999  
 C:Accession: A01244  
 R:Heilng, R.; Muraskowsky, R.; Kloeffer, C.; Mandel, J.L.  
 Nucleic Acids Res. 10, 4363-4382, 1982  
 A:Title: The ovalbumin gene family: complete sequence and structure of the y gene.  
 A:Reference number: A01244; MUID:83014329; PMID:7122240  
 A:Accession: A01244  
 A:Molecule type: DNA  
 A:Residues: 1-388 <HEI>  
 A:Cross-references: GB:J00922; GB:V00439; NID:9212899; PIDN:AAA68882.1; PID:9212900  
 C:Genetics:  
 A:Introns: 56/3; 73/3; 116/3; 156/1; 203/3; 255/3  
 C:Superfamily: antithrombin III  
 C:Keywords: glycoprotein; phosphoprotein; serine proteinase inhibitor  
 F:74-121/disulfide bonds: #status predicted  
 F:293/Binding site: carboxylate (Asn) (covalent) #status predicted  
 F:345/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 72.3%; Score 60; DB 1; Length 388;  
 Best Local Similarity 68.8%; Pred. No. 0.034;  
 Matches 11; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 2 YIEVTEGTEATATG 17  
 ::| | | | | | | |  
 Db 332 FMEVNEGTEATGSGT 347

RESULT 6  
 DXBH2  
 protein 24 - barley  
 N:Alternate names: major endosperm albumin  
 C:Species: Hordeum vulgare (barley)  
 C:Date: 30-Jun-1987 #sequence\_revision 01-Dec-2000 #text\_change 01-Dec-2000  
 C:Accession: S13822; A01252  
 R:Brandt, A.; Svendsen, I.; Hejgaard, J.  
 Eur. J. Biochem. 194, 499-505, 1990  
 A:Title: A plant serpin gene, structure, organization and expression of the gene enco  
 A:Reference number: S13822; MUID:91099324; PMID:2269280  
 A:Accession: S13822  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-399 <BRA>  
 A:Cross-references: GB:X51726; NID:q19067; PIDN:CA36015.1; PID:q19068  
 R:Hejgaard, J.; Rasmussen, S.K.; Brandt, A.; Svendsen, I.  
 FEBS Lett. 180, 89-94, 1985  
 A:Title: Sequence homology between barley endosperm protein 2 and protease inhibitors



A:Reference number: A91332  
A:Accession: A01252  
A:Molecule type: mRNA  
A:Residues: 220-399 <HEJ>  
A:Cross-references: GB:X05902; NID:g19078; PIDN:CAA29331.1; PID:g19079  
A:Note: most of the sequence has been confirmed by amino acid sequencing  
R:Nielsen, G.; Johansen, H.; Jensen, J.; Hejgaard, J.  
Barley Genet. Newsl. 13, 55-57, 1983  
A:Title: Localization on barley chromosome 4 of genes coding for beta-amylase (Amy1) and  
A:Reference number: A90095  
A:Accession: A90095  
A:Contents: annotation: map position  
C:Comment: A major component of the endosperm albumin, this protein acts as a storage pr  
ins of the antithrombin-III superfamily suggests that it also has an inhibitory function  
C:Genetics:  
A:Gene: Pz1  
A:Map position: 4  
C:Superfamily: antithrombin III  
C:Keywords: serine proteinase inhibitor; storage protein  
F:357/inhibitory site: Met (unidentified proteinase) #status predicted

Query Match 72.38; Score 60; DB 1; Length 399;  
Best Local Similarity 75.08; Pred. No. 0.035;  
Matches 12; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 335 SFEVNEEGTEAGNAT 350

RESULT 7  
A57488  
A:Species: Mus musculus (house mouse)  
C:Date: 08-Dec-1995 #sequence\_revision 08-Dec-1995 #text\_change 16-Jul-1999  
C:Accession: A57488  
R:Sun, J.; Rose, J.B.; Bird, P.  
J. Biol. Chem. 270, 16089-16096, 1995  
A:Title: Gene structure, chromosomal localization, and expression of the murine homology  
A:Reference number: A57488; MUID:95332310; PMID:7608171  
A:Accession: A57488  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-378 <SUN>  
A:Cross-references: GB:U25844; NID:g818902; PIDN:AAA79684.1; PID:g818903  
C:Genetics:  
A:Gene: Sp13  
A:Map position: 13  
C:Superfamily: antithrombin III  
C:Keywords: serine proteinase inhibitor  
F:343/inhibitory site: Arg (unidentified proteinase) #status predicted

Query Match 71.18; Score 59; DB 2; Length 378;  
Best Local Similarity 68.88; Pred. No. 0.049;  
Matches 11; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 321 AFVEVNEEGTEAANAAT 336

RESULT 8  
I38201  
A:Species: Homo sapiens (man)  
C:Date: 23-Feb-1996 #sequence\_revision 23-Feb-1996 #text\_change 26-May-2000  
C:Accession: I38201; G01631; JT0966; JT0967  
R:Schneider, S.S.; Schick, C.; Fish, K.E.; Miller, E.; Pena, J.C.; Treter, S.D.; Hul, S.  
Proc. Natl. Acad. Sci. U.S.A. 92, 3147-3151, 1995  
A:Title: A serine proteinase inhibitor locus at 18q21.3 contains a tandem duplication of  
A:Reference number: I38200; MUID:95241462; PMID:7724531  
A:Accession: I38201  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA

A:Residues: 1-390 <SCH1>  
A:Cross-references: EMBL:U19568; GB:U19558; NID:g1172085  
A:Accession: I38200  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-117 <SCH2>  
A:Cross-references: EMBL:U19562; NID:g897835; PIDN:AAA6316.1; PID:g897844  
R:Silvman, G.A.  
submitted to the EMBL Data Library, January 1995  
A:Reference number: G07968  
A:Accession: G01631  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-350, 'G', 352-390 <STL>  
A:Cross-references: EMBL:U19556; NID:g1276435; PID:g1052869  
R:Suminami, Y.; Kishi, F.; Sekiguchi, K.; Kato, H.  
Biochem. Biophys. Res. Commun. 181, 51-58, 1991  
A:Title: Squamous cell carcinoma antigen is a new member of the serine protease inhib  
A:Reference number: JT0966; MUID:92068241; PMID:1958219  
A:Accession: JT0966

A:Molecule type: mRNA  
A:Residues: 1-350, 'G', 352-356, 'A', 358-390 <SUM1>  
A:Cross-references: GB:S66896; NID:g239551; PIDN:AAB20405.1; PID:g239552  
A:Accession: JT0967

A:Molecule type: protein  
A:Residues: 11-21, 231-237, 240-256, 303-325 <SUM2>  
C:Comment: This antigen probably acts as a proteinase inhibitor to modulate the host  
C:Genetics:  
A:Gene: GDB:SCCA1; SCC  
A:Cross-references: GDB:625364; OMIM:600517  
A:Map position: 18q21.3-18q21.3  
A:Residues: 55/3; 74/3; 117/3; 157/1; 204/3; 256/3  
C:Superfamily: antithrombin III  
C:Keywords: cysteine proteinase inhibitor; glycoprotein  
F:65,93,171,376/Binding site: carboxydrate (Asn) (covalent) #status predicted  
F:354/inhibitory site: Ser (cathepsin L) #status predicted

Query Match 71.18; Score 59; DB 2; Length 390;  
Best Local Similarity 68.88; Pred. No. 0.05;  
Matches 11; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 332 AFVEVTEGAEANAAT 347

RESULT 9  
I38202  
A:Species: Homo sapiens (man)  
C:Date: 23-Feb-1996 #sequence\_revision 23-Feb-1996 #text\_change 26-May-2000  
C:Accession: I38202; S66675; S57522  
R:Schneider, S.S.; Schick, C.; Fish, K.E.; Miller, E.; Pena, J.C.; Treter, S.D.; Hul, S.  
Proc. Natl. Acad. Sci. U.S.A. 92, 3147-3151, 1995  
A:Title: A serine proteinase inhibitor locus at 18q21.3 contains a tandem duplication  
A:Reference number: I38200; MUID:95241462; PMID:7724531  
A:Accession: I38202  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-390 <SCH>  
A:Cross-references: EMBL:U19576; GB:U19569; NID:g852466  
R:Barnes, R.C.; Morrall, D.M.  
FEBS Lett. 373, 61-65, 1995  
A:Title: Identification of a novel human serpin gene; cloning sequencing and expressi  
A:Reference number: S66675; MUID:96013887; PMID:7589435  
A:Accession: S66675  
A:Molecule type: mRNA  
A:Residues: 7-351, 'V', 353-384 <BAR>  
A:Cross-references: EMBL:X85015; NID:g887464; PIDN:CAA61420.1; PID:g887465  
C:Genetics:  
A:Gene: GDB:SCCA2; P11  
A:Cross-references: GDB:636556; OMIM:600518

A:Map position: 18q21.3-18q21.3  
 A:introns: 55/3; 74/3; 117/3; 157/1; 204/3; 256/3  
 C:Superfamily: antithrombin III  
 C:Keywords: glycoprotein; serine proteinase inhibitor  
 F:1-22/Domain: signal sequence #status predicted <SIG>  
 F:23-384/Product: leupin #status predicted <MAT>  
 F:65,93,170,376/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F:354/Inhibitory site: Leu (unidentified proteinase) #status predicted

Query Match 71.1%; Score 59; DB 2; Length 390;  
 Best Local Similarity 68.8%; Pred. No. 0.05;  
 Matches 11; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 SYIEVTEGTEATAT 16  
 DB 332 AFVEVTEGVEAAAT 347  
 ::::| | | | | | | | | | | |

RESULT 10  
 A59273  
 Proteinase inhibitor 8 - human  
 M:Alternate names: cytoplasmic antiproteinase 2 (CAP-2)  
 C:Species: Homo sapiens (man)  
 C>Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000  
 C:Accession: A59273  
 R:Spracher, C.A.; Morgenstern, K.A.; Mathewes, S.; Dahlen, J.R.; Schrader, S.K.; Foster, J. Biol. Chem. 270, 29854-29861, 1995  
 A:Title: Molecular cloning, expression, and partial characterization of two novel member  
 A:Reference number: A59273; MUID:96102039; PMID:8530382  
 A:Accession: A59273  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-374 <SPR>  
 A:Cross-references: GB:LA0377; NID:q1160926; PIDN:AA041939.1; PID:q1160927  
 C:Experimental source: tissue type Placenta; note (vector lambda gtl1); gene CAP2  
 C:Genetics:

A:Gene: GDB:PI8; CAP2  
 A:Cross-references: GDB:599392; OMIM:601697  
 A:Map position: 18q21.3-18q21.3  
 C:Superfamily: antithrombin III  
 C:Keywords: proteinase inhibitor  
 F:339/Inhibitory site: Arg (unidentified proteinase) #status predicted

Query Match 69.9%; Score 58; DB 2; Length 374;  
 Best Local Similarity 73.3%; Pred. No. 0.071;  
 Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 YIEVTEGTEATAT 16  
 DB 318 FEVNEEGTEAAAT 332  
 ::::| | | | | | | | | | | |

RESULT 11  
 B59273  
 Proteinase inhibitor 9 - human  
 M:Alternate names: cytoplasmic antiproteinase 3 (CAP-3)  
 C:Species: Homo sapiens (man)  
 C>Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000  
 C:Accession: B59273  
 R:Spracher, C.A.; Morgenstern, K.A.; Mathewes, S.; Dahlen, J.R.; Schrader, S.K.; Foster, J. Biol. Chem. 270, 29854-29861, 1995  
 A:Title: Molecular cloning, expression, and partial characterization of two novel member  
 A:Reference number: A59273; MUID:96102039; PMID:8530382  
 A:Accession: B59273  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-376 <SPR>  
 A:Cross-references: GB:LA0378; NID:q1160928; PIDN:AA041940.1; PID:q1160929  
 C:Experimental source: tissue type Placenta; note (vector lambda gtl1); gene CAP2; tissue  
 C:Genetics:

A:Gene: GDB:PI9; CAP3  
 A:Cross-references: GDB:599393; OMIM:601799  
 A:Map position: 6p25-6p25

C:Superfamily: antithrombin III  
 C:Keywords: proteinase inhibitor  
 F:340/Inhibitory site: Glu (unidentified proteinase) #status predicted

Query Match 69.9%; Score 58; DB 2; Length 376;  
 Best Local Similarity 68.8%; Pred. No. 0.071;  
 Matches 11; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 SYIEVTEGTEATAT 16  
 DB 318 FEVNEEGTEAAAT 332  
 ::::| | | | | | | | | | | |

RESULT 12  
 T06183  
 serpin - barley  
 C:Species: Hordeum vulgare (barley)  
 C>Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 21-Jul-2000  
 C:Accession: T06183  
 R:Rasmussen, S.K.; Klausen, J.; Hejgaard, J.; Svenson, B.; Svendsen, I. Biochim. Biophys. Acta 1297, 127-130, 1996  
 A:Title: Primary structure of the plant serpin B57 having the capacity of chymotryps  
 A:Reference number: 215515; MUID:97075187; PMID:8917613  
 A:Accession: T06183  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-397 <RAS>  
 A:Cross-references: EMBL:X95277; NID:q1197576; PIDN:CAA64599.1; PID:q1197577  
 A:Experimental source: cv. Bomi, endosperm  
 C:Superfamily: antithrombin III

Query Match 69.9%; Score 58; DB 2; Length 397;  
 Best Local Similarity 64.7%; Pred. No. 0.075;  
 Matches 11; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 SYIEVTEGTEATATG 17  
 DB 336 FEVDEEGTKAGATG 352  
 ::::| | | | | | | | | | | |

RESULT 13  
 AH1903  
 hypothetical protein all0778 [imported] - Nostoc sp. (strain PCC 7120)  
 C:Species: Nostoc sp.  
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C>Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
 C:Accession: AH1903  
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigu  
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata  
 DNA Res. 8, 205-213, 2001  
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium  
 A:Reference number: AB1807; MUID:21595285; PMID:11759840  
 A:Accession: AH1903  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-374 <KUR>  
 A:Cross-references: GB:BA000019; PIDN:BA072735.1; PID:q17130123; GSPDB:GN00179  
 C:Experimental source: strain PCC 7120  
 C:Genetics:

A:Gene: all0778

Query Match 68.7%; Score 57; DB 2; Length 374;  
 Best Local Similarity 85.7%; Pred. No. 0.1;  
 Matches 12; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 IEVTEGTEATAT 16  
 DB 315 IEVNEEGTEASAT 328  
 ::::| | | | | | | | | | | |

RESULT 14  
 S29819  
 serpin - barley

C:Species: Hordeum vulgare (barley)  
 C:Date: 02-Dec-1993 #sequence\_revision 12-Apr-1996 #text\_change 16-Jul-1999  
 C:Accession: S29819; S74166; S25625  
 R:Rasmussen, S.K.  
 Blochlm. Biophys. Acta 1172, 151-154, 1993  
 A:Title: A gene coding for a new plant serpin.  
 A:Reference number: S29819; MUID:93176798; PMID:8439552  
 A:Accession: S29819  
 A:Molecule type: DNA  
 A:Residues: 1-398 <RAS>  
 A:Cross-references: EMBL:Z15116; NID:919070; PIDN:CAA78822.1; PID:919071  
 A:Experimental source: strain Villa x Ulla  
 R:Dahl, S.W.; Rasmussen, S.K.; Petersen, L.C.; Hejgaard, J.  
 FEBS Lett. 394, 165-168, 1996  
 A:Title: Inhibition of coagulation factors by recombinant barley serpin BSZx.  
 A:Reference number: S74166; MUID:97000013; PMID:8843156  
 A:Accession: S74166  
 A:Molecule type: protein  
 A:Residues: 357-361 <DAH>  
 C:Genetics:  
 A:Gene: pazx  
 A:introns: 136/3  
 C:Superfamily: antithrombin III

Query Match 68.7%; Score 57; DB 2; Length 398;  
 Best Local Similarity 68.8%; Pred. No. 0.11;  
 Matches 11; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 SYEVTEGTEATAT 16  
 I::I I I I I I I I  
 Db 334 SEVEVNEGTEAAART 349

RESULT 15  
 A26423  
 serine protease inhibitor 2.1 - rat (fragment)  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 19-Nov-1988 #sequence\_revision 12-Jul-1996 #text\_change 12-Jul-1996  
 C:Accession: A26423  
 R:Hill, R.E.; Hastie, N.D.  
 Nature 326, 96-99, 1987  
 A:Title: Accelerated evolution in the reactive centre regions of serine protease inhibitor  
 A:Reference number: A26423; MUID:87144617; PMID:3493437  
 A:Accession: A26423  
 A:Molecule type: mRNA  
 A:Residues: 1-213 <HIL>  
 A:Cross-references: GB:X13149  
 A:Note: the authors translated the codon CTG for residue 1 as Met, CTG for residue 35 as  
 11 as Met  
 C:Superfamily: antithrombin III

Query Match 67.5%; Score 56; DB 2; Length 213;  
 Best Local Similarity 73.3%; Pred. No. 0.084;  
 Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 3 IEVTEGTEATATG 17  
 ::I I I I I I I I I I  
 Db 159 LDVAETGTETATATG 173

Search completed: July 11, 2003, 12:00:36  
 Job time : 3.52827 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:43:41 ; Search time 1.62578 seconds  
(without alignments)  
433.698 Million cell updates/sec

Title: US-10-091-442-4

Perfect score: 83

Sequence: 1 SYIEVTEECTEATATG 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	83	100.0	380	1 SPB7_HUMAN	O75635 homo sapien
2	73	88.0	378	1 PT16_BOVIN	O02739 bos taurus
3	69	83.1	392	1 SB11_HUMAN	O96p15 homo sapien
4	67	80.7	391	1 SB13_HUMAN	O9u18 homo sapien
5	62	74.7	376	1 PT16_HUMAN	P35237 homo sapien
6	62	74.7	379	1 PT16_HUMAN	P30740 homo sapien
7	60	72.3	388	1 OVAY_CHICK	P01014 gallus gall
8	60	72.3	399	1 PT16_HUMAN	P06293 hordeum vul
9	59	71.1	378	1 PT16_HUMAN	P80229 sus scrofa
10	59	71.1	380	1 PT16_MOUSE	O60854 mus musculu
11	59	71.1	390	1 SCCL_HUMAN	P29508 homo sapien
12	59	71.1	390	1 SCCL_HUMAN	P48554 homo sapien
13	58	69.9	374	1 SPB8_HUMAN	P50452 homo sapien
14	58	69.9	376	1 SPB8_HUMAN	P50452 homo sapien
15	57	68.7	410	1 NEUS_RAT	O91142 rattus norv
16	57	68.7	214	1 SI21_RAT	P09005 rattus norv
17	56	67.5	379	1 TLEU_HORSE	P05619 equus cabal
18	56	67.5	405	1 SB12_HUMAN	O96p13 homo sapien
19	56	67.5	410	1 NEUS_CHICK	O90935 gallus gall
20	56	67.5	410	1 NEUS_MOUSE	O35664 mus musculu
21	56	67.5	418	1 SI24_APOSY	O60336 apodemus sy
22	55	66.3	410	1 NEUS_HUMAN	O99574 homo sapien
23	55	66.3	413	1 CP13_RAT	P05544 rattus norv
24	54	65.1	415	1 PA12_HUMAN	P05120 homo sapien
25	53	63.9	397	1 SB10_HUMAN	P48595 homo sapien
26	52	62.7	222	1 SPIB_VACCC	P20842 vaccinia vi
27	52	62.7	341	1 SPI2_COMPR	P07385 compox viru
28	52	62.7	344	1 SPI2_VARV	P33830 variola vir
29	52	62.7	345	1 SPI2_RABPU	P42926 rabbitpox v
30	52	62.7	345	1 SPI2_VACCV	P15059 vaccinia vi
31	51	61.4	353	1 SPI1_VACCC	P20531 vaccinia vi
32	51	61.4	353	1 SPI1_VACCV	P15058 vaccinia vi
33	51	61.4	357	1 SPI1_RABPU	P42928 rabbitpox v

34	51	61.4	397	1 GDN_MOUSE	O07235 mus musculu
35	51	61.4	398	1 GDN_HUMAN	P07093 homo sapien
36	51	61.4	418	1 COTR_MOUSE	P07759 mus musculu
37	51	61.4	418	1 CP16_RAT	P09006 rattus norv
38	51	61.4	433	1 ANT3_BOVIN	P41361 bos taurus
39	51	61.4	464	1 ANT3_HUMAN	P01008 homo sapien
40	51	61.4	465	1 ANT3_MOUSE	P32261 mus musculu
41	51	61.4	465	1 ANT3_SHEEP	P32262 ovis aries
42	50	60.2	355	1 SPI1_COMPR	P42927 compox viru
43	50	60.2	357	1 SPI1_VARV	P33829 variola vir
44	50	60.2	392	1 SERA_MANSE	P14754 manduca sex
45	50	60.2	405	1 SPI2_MOUSE	O9jx88 mus musculu

## ALIGNMENTS

RESULT 1  
ID SPB7\_HUMAN STANDARD; PRT; 380 AA.  
AC O75635;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Mgsin (TP55) (Serpin B7).  
GN SERPINB7  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-97326116; PubMed-9182567;  
RA Tsujimoto M., Tsuturo K., Ishida N., Kurihara T., Iwasa F.,  
Yamashiro K., Rogi T., Kodama S., Katsuragi N., Adachi M.,  
Katayama T., Nakao M., Yamachi K., Hashino J., Haruyama M., Mura K.,  
Nakanishi T., Nakazato H., Teramura M., Mizoguchi H., Yamaguchi N.;  
RT "Purification, cDNA cloning, and characterization of a new serpin with  
megakaryocyte maturation activity.";  
RL J. Biol. Chem. 272:15373-15380(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-98376492; PubMed-9710452;  
RA Miyata T., Nangaku M., Suzuki D., Inagi R., Uragami K., Sakai H.,  
Okubo K., Kurokawa K.;  
RT "A mesangium-predominant gene, mgsin, is a new serpin upregulated in  
IgA nephropathy.";  
RL J. Clin. Invest. 102:828-836(1998).  
CC -!- FUNCTION: Might influence the maturation of Lys-specific  
proteases. Might influence the maturation of megakaryocytes via  
its action as a serpin.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -!- TISSUE SPECIFICITY: Predominantly expressed in mesangial cells.  
CC -!- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
CC EMBL: D88575; BAA31232.1; -  
CC EMBL: AF027866; AAC64506.1; -  
CC HSP: P05619; IHL.  
CC HSP: HGNC:13902; SERPINB7.  
CC MTM: 603357; -  
CC InterPro: IPR000215; Serpin.  
CC Pfam: PF00079; serpin; 1.  
CC SMART: SM00093; SERPIN; 1.  
CC PROSITE: PS00284; SERPIN; 1.

KW Serpin; Serine protease inhibitor.  
 FT ACT\_SITE 347 REACTIVE BOND (BY SIMILARITY).  
 SQ SEQUENCE 380 AA; 42904 MW; 9A2CDB6C63CFF605 CRC64;

Query Match 100.0%; Score 83; DB 1; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-06;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIEVTEGTEATATG 17  
 DB 325 SYIEVTEGTEATATG 341

## RESULT 2

PT16\_BOVIN STANDARD; PRT; 378 AA.  
 AC 002739;  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Serine proteinase inhibitor B-43.  
 OS Bos taurus (Bovine).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 CC Bovidae; Bovinae; Bos.  
 OX NCBI\_Taxid=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=97165893; PubMed=9013786;  
 RA Nakaya N., Nishibori M., Kawabata M., Saeki K.;  
 RT "Cloning of a serine proteinase inhibitor from bovine brain:  
 RT expression in the brain and characterization of its target  
 RT proteinases.";  
 RL Brain Res. Mol. Brain Res. 42:293-300(1996).  
 CC PRESENT IN THE BRAIN OR EXTRAVASATED FROM THE BLOOD.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- TISSUE SPECIFICITY: BRAIN.  
 CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: D55670; BAA19875.1; -  
 DR HSSP: P05120; 1BY7.  
 DR InterPro: IPR000215; Serpin.  
 DR Pfam: PF00079; serpin.1.  
 DR SMART: SM00093; SERPIN.1.  
 DR PROSITE: PS00284; SERPIN.1.  
 DR Serpin; Serine protease inhibitor.  
 KW ACT\_SITE 343 344 REACTIVE BOND (BY SIMILARITY).  
 FT ACT\_SITE 343 344  
 SQ SEQUENCE 378 AA; 42560 MW; 664F49CCCE263A CRC64;

Query Match 88.0%; Score 73; DB 1; Length 378;  
 Best Local Similarity 82.4%; Pred. No. 6.2e-05;  
 Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 SYIEVTEGTEATATG 17  
 DB 321 SFVEVTEGTEAATG 337

## RESULT 3

SB11\_HUMAN STANDARD; PRT; 392 AA.  
 AC Q96P15; Q96P14; Q96P13;  
 DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Serpin B11.  
 GN SERPINB11.

OS Homo sapiens (Human)  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND VARIANTS THR-148 AND THR-181.  
 RA Silverman G.A., Cataltepe S.;  
 RT "SERPINB11 is a novel ov-serpin with at least 3 alleles.";  
 RT Submitted (SEP-2001) to the EMBL/Genbank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----

DR EMBL: AF419953; AAL16056.1; -  
 DR EMBL: AF419954; AAL16057.1; -  
 DR EMBL: AF419955; AAL16058.1; -  
 DR Genew: HGNC:14221; SERPINB11.  
 DR InterPro: IPR000215; Serpin.  
 DR Pfam: PF00079; serpin.1.  
 DR PROSITE: PS00284; SERPIN.1.  
 KW Serpin; Serine protease inhibitor; Polymorphism.  
 FT ACT\_SITE 357 358 REACTIVE BOND (BY SIMILARITY).  
 FT ACT\_SITE 148 148 M -> T (IN ALLELE B).  
 FT VARIANT 148 148 /FTIG-Var\_012472  
 FT VARIANT 181 181 I -> T (IN ALLELE B AND ALLELE C).  
 FT FTIG-Var\_012473.  
 SQ SEQUENCE 392 AA; 44098 MW; 906F6DD412BCD756 CRC64;

Query Match 83.1%; Score 69; DB 1; Length 392;  
 Best Local Similarity 76.5%; Pred. No. 0.00031;  
 Matches 13; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 SYIEVTEGTEATATG 17  
 DB 335 SYLDVSEGTETAAATG 351

## RESULT 4

SB13\_HUMAN STANDARD; PRT; 391 AA.  
 AC Q9U1V8; Q9UBW1; Q9UGK0; Q9HCX1.  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hurpin (Hcar UV-repressible serpin) (Protease inhibitor 13)  
 DE (Headpin) (Serpin B13).  
 GN SERPINB13 OR P113.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC TISSUE=Skin;  
 RX MEDLINE=99443898; PubMed=10512713;  
 RA Abts H.F., Welts T., Mirmohammadsadeh A., Koehrer K., Michel G.,  
 RA Ruzicka T.;  
 RT "Cloning and characterization of hurpin (Protease inhibitor 13): a new  
 RT skin specific, UV-repressible serine proteinase inhibitor of the  
 RT ovalbumin serpin family.";  
 RL J. Mol. Biol. 293:29-39(1999).  
 RN [2]

RP SEQUENCE FROM N.A. (ISOFORM 1).  
RX MEDLINE-99458661; PubMed-10527891;  
RA Spring P., Nakashima T., Frederick M., Henderson Y., Clayman G.,  
RT "Identification and cDNA cloning of headpin, a novel differentially  
RL Blochem. Biophys. Res. Commun. 264:299-304(1999).  
RN [3]  
RN SEQUENCE FROM N.A. (ISOFORM 1).  
RX MEDLINE-20461850; PubMed-11004515;  
RA Nakashima T., Pak S.C., Silverman G.A., Spring P.M., Frederick M.J.,  
RT Clayman G.L.;  
RT "Genomic cloning, mapping, structure and promoter analysis of HEADPIN,  
RL a serpin which is down-regulated in head and neck cancer cells.";  
RN Blochim. Biophys. Acta 1492:441-446(2000).  
[4]  
RP SEQUENCE FROM N.A. (ISOFORM 2).  
RX TISSUE-Placenta;  
RX MEDLINE-21213570; PubMed-11313015;  
RA Abbs H.F., Weiss T., Scheuring S., Scott F.L., Irving J.A., Michel G.,  
RT Bird P.I., Ruzicka T.;  
RT "Sequence, organization, chromosomal localization and alternative  
RT splicing of the human serine protease inhibitor gene hurpin (P113),  
RL DNA Cell Biol. 20:123-131(2001).  
CC -1- FUNCTION: May play a role in the proliferation or differentiation  
CC of keratinocytes.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms: 1 (shown here) and 2; are  
CC produced by alternative splicing.  
CC -1- TISSUE SPECIFICITY: SKIN-SPECIFIC.  
CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL: AJ001696; CAA04935.2; -  
DR EMBL: AJ001697; CAA04936.2; -  
DR EMBL: AJ001698; CAA04937.1; -  
DR EMBL: AF169949; AAD5765.1; -  
DR EMBL: AF216854; AAF72879.1; -  
DR EMBL: AJ278717; CAC03569.1; -  
DR HSSP: P05120; 1BY7.  
DR Genew: HGNC:8944; SERPINB13.  
DR MIM: 604445; -  
DR InterPro: IPR000240; Maspin.  
DR InterPro: IPR000215; Serpin.  
DR Pfam: PF00079; serpin. 1.  
DR PRINTS: PR00676; MASPIN.  
DR SMART: SM00093; SERPIN. 1.  
DR PROSITE: PS00284; SERPIN. 1.  
RW Serpin: Serine protease inhibitor; Alternative splicing.  
FT ACT\_SITE 356 357 REACTIVE\_BOND (BY SIMILARITY).  
FT VAAPSLIC 206 257 MISSING (IN ISOFORM 2).  
FT CONFLICT 8 8 S -> N (IN REF. 4).  
FT CONFLICT 75 75 MISSING (IN REF. 1; CAA04937).  
FT CONFLICT 293 293 G -> S (IN REF. 2).  
FT CONFLICT 297 297 E -> Q (IN REF. 1; CAA04937).  
SO SEQUENCE 391 AA; 44376 MW; 2CA88558D4BC2B09 CRC64;  
Query Match 80.7%; Score 67; DB 1; Length 391;  
Best Local Similarity 76.5%; Pred. No. 0.00067;  
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```
Qy      1 SYIEVTEEGTEATAATG 17
        |::| | | | | | | |
Db      334 SFVAVTEEGTEAANAATG 350
```

```

RESULT 5
PT16_HUMAN
ID ID PT16_HUMAN STANDARD: PRT: 376 AA.
AC P35237; Q96U44;
DT 01-FEB-1994 (Rel. 28, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Placental thrombin inhibitor (Cytoplasmic antiproteinase) (CAP)
GN SERPINB6 OR PI6 OR PTI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_Taxid:9606;
RN [1]
RP
RC TISSUE=Placenta;
RC MEDLINE=94022386; PubMed=8415716;
RA Coughlin P., Sun J., Cerruti L., Salem H.H., Bird P.;
RT "Cloning and molecular characterization of a human intracellular
RT serine proteinase inhibitor."
RT Proc. Natl. Acad. Sci. U.S.A. 90:9417-9421(1993).
RL [2]
RN
RP
RC TISSUE=Placenta; AND SEQUENCE OF 67-73 AND 144-149.
RC MEDLINE=94183847; PubMed=8136380;
RX Morgenstern K.A., Sprecher C.A., Holth L., Foster D., Grant F.J.,
RA Ching A., Kistiel W.;
RT "Complementary DNA cloning and kinetic characterization of a novel
RT intracellular serine proteinase inhibitor: mechanism of action with
RT trypsin and factor Xa as model proteinases."
RL Biochemistry 33:3432-3441(1994).
RN [3]
RP
RC TISSUE=Placenta;
RC TISSUE=Colon;
RC Strausberg R.;
RL submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Inhibits thrombin.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: HIGHEST LEVELS IN SKELETAL MUSCLE. ALSO FOUND
CC IN PLACENTA; CARDIAC MUSCLE, LUNG, LIVER, KIDNEY AND PANCREAS.
CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC The European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/sib-ch).
CC -----
DR EMBL; Z22658; CAAB0373.1; -
DR EMBL; S692722; AAB30320.1; -
DR EMBL; BC001394; AAH01394.1; -
DR PIR; S35750; S35750.
DR PIR; A48681; A48681.
DR HSSP; P05120; 1BY7.
DR Genew; HGNC:8950; SERPINB6.
DR MIM; 173321; -
DR InterPro; IPR000215; Serpin.
DR Pfam; PF00079; serpin; 1.
DR SMART; SM00093; SERPIN; 1.
DR PROSITE; PS00284; SERPIN; 1.
KW Serpin; Serine protease Inhibitor.
FT ACT_SITE 341 342 REACTIVE BOND.
FT CONFLICT 175 175 E -> G (IN REF. 1).
FT CONFLICT 362 362 S -> R (IN REF. 1).
SQ SEQUENCE 376 AA; 42590 MW; 2B46A5F40C60806C CRC64;

```

Query Match	74.7%;	Score 62;	DB 1;	Length 376;
Best Local Similarity	75.0%;	Pred. No. 0.0045;		
Matches 12;	Conservative 2;	Mismatches 2;	Indels 0;	Gaps 0;





```

DT      15-JUN-2002 (Rel..41, Last annotation update)
DE      Protein Z (Z4) (Major endosperm albumin).
CN      PAZ1.
OC      Hordeum vulgare (Barley).
OC      Euxaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophytes; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
OC      Triticeae; Hordeum.
OX      NCBI_TaxId=4513;
RN      [1]
RN      SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RP      STRAIN=cv. Carlberg II.; TISSUE=grain;
RC      MEDLINE=9109324; Pubmed=2269280;
RA      Hejgaard J., Rasmussen S.K., Brandt A., Svendsen I.;
RT      "Sequence homology between barley endosperm protein Z and protease
RL      inhibitors of the alpha-1-antitrypsin family.";
RN      FEBS Lett. 180:89-94(1985).
RN      [3]
RN      SEQUENCE OF 360-384 FROM N.A.
RP      Rasmussen S.K., Hopp H.E., Brandt A., Svendsen I., Hejgaard J.;
RA      "A cDNA clone for protein Z, a major barley endosperm albumin.";
RT      Carlberg Res. Commun. 49:385-390(1984).
RN      RL
CC      -1- FUNCTION: A MAJOR COMPONENT OF THE ENDOSPERM ALBUMIN, THIS PROTEIN
CC      ACTS AS A STORAGE PROTEIN DURING GRAIN FILLING, CONTRIBUTING A
CC      SUBSTANTIAL PART OF THE GRAIN'S LYSINE. MAY HAVE AN INHIBITORY
CC      FUNCTION DURING FILLING OR GERMINATION.
CC      -1- TISSUE SPECIFICITY: IS ACCUMULATED AND STORED IN THE ENDOSPERM,
CC      WHERE IT EXISTS IN A FREE AND A BOUND FORM.
CC      -1- DEVELOPMENTAL STAGE: SYNTHESIZED 10-25 DAYS AFTER FERTILIZATION
CC      (DEVELOPING ENOSPERM).
CC      -1- INDUCTION: ITS EXPRESSION IS REGULATED BY THE "HIGH LYSINE"
CC      ALLIELES LYSI AND LYS3A.
CC      CC      MISCELLANEOUS: THERE SEEM TO BE TWO Z PROTEINS: Z4 (FROM
CC      CHROMOSOME 4) AND Z7 (FROM CHROMOSOME 7).
CC      -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outpost -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X51726; CAA36015.1; .
DR      EMBL; X05902; CAA29331.1; .
DR      EMBL; M35065; AAA32971.1; -.
DR      PIR; A01252; DXBHZ.
DR      PIR; S13822; S13822.
DR      HSSP; P01008; 1ATH.
DR      InterPro; IPRO00215; Serpin.
DR      Pfam; PF00079; serpin. 1.
DR      SMART; SM00093; SERPIN. 1.
DR      ProSITE; PS00284; SERPIN. 1.
KW      Plant defense; Serpin; Storage protein; Multigene family.
FT      DOMAIN
FT      36
FT      56
FT      FT      SIGNAL FOR TARGETING PROTEIN Z4 INTO
FT      THE ER LUMEN (POTENTIAL).
FT      REACTIVE BOND (POTENTIAL).
SO      ACCT SITE    357       357
SEQUENCE 399 AA: 43276 MW: 023B8E52475BA56 CRC64;
```

ID	FILEU_PIG	STANDARD:	PRT:	378 AA.
AC	P80229;			
DT	01-OCT-1993 (Rel. 27, Created)			
DT	01-OCT-1993 (Rel. 27, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Leukocyte elastase inhibitor (LEI) (leucocyte neutral proteinase inhibitor) (PI-6).			
GN	SERPINF1 OR ELANH2.			
OS	Sus scrofa (Pig).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.			
OX	NCBI_taxid:9823;			
RN	[1]			
RP	SEQUENCE.			
RC	TISSUE=Leukocyte;			
RC	MEDLINE=94039085; PubMed=7901009;			
RA	Teschauer W.F., Mentele R., Sommerhoff C.P.;			
RT	"Primary structure of a porcine leukocyte serpin."			
RL	Eur. J. Biochem. 217:519-526(1993).			
CC	-1- FUNCTION: REGULATES THE ACTIVITY OF THE NEUTROPHIL PROTEASE ELASTASE. CAPREPSIN G AND PROTEINASE-3 (By similarity).			
CC	-1- SUBUNIT: HOMODIMER; DISULFIDE-LINKED.			
CC	-1- SUBCELLULAR LOCATION: Cytoplasmic.			
DR	HSSP; P05619; 1HLE.			
DR	InterPro: IPR000215; Serpin.			
DR	SMART; SM00093; SERPIN; 1.			
DR	PROSITE; PS00284; SERPIN; 1.			
KW	Serpin; Serine protease inhibitor.			
FT	MOD_RES 1	BLOCKED.		
FT	DISULFID 80	INTERCHAIN (PROBABLE).		
FT	ACT SITE 343 344	REACTIVE BOND (BY SIMILARITY).		
SQ	SEQUENCE 378 AA: 42512 MW; 3CA7C65F11E508CA CRC64;			
OY	1 SYIETECTETATAT 16			
DB	321 AFVEYNEGTETAAAT 336			
DT	15-JUN-2002 (Rel. 41, Created)			
DT	15-JUN-2002 (Rel. 41, Last sequence update)			
DE	Placental thrombin inhibitor (Protease inhibitor 6) (PI-6).			
GN	SERPINF6 OR SPI3			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.			
OX	NCBI_taxid:10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BALB/c;			
RC	MEDLINE=95332310; PubMed=7608171;			
RA	Sun J., Rose J.B., Bird P.;			
RT	"Gene structure, chromosomal localization, and expression of the murine homologue of human proteinase inhibitor 6 (PI-6) suggests divergence of PI-6 from the ovalbumin serpins."			
RL	J. Biol. Chem. 270:16089-16096(1995).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Breast tumor;			

RA Strausberg R.;  
 RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Inhibits thrombin.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U25844; AAA79684.1; -  
 DR EMBL: BC006766; AAH06766.1; -  
 DR HSSP: P05120; 1BY7.  
 DR MGD: MGI:103123; Serpinb6.  
 DR InterPro: IPR000215; Serpin.  
 DR Pfam: PF00079; serpin; 1.  
 DR SMART: SM00093; SERPIN; 1.  
 DR PROSITE: PS00284; SERPIN; 1.  
 KW Serpin; Serine protease inhibitor.  
 FT ACT\_SITE 343 344  
 SQ SEQUENCE 378 AA; 42599 MW; 4B0F5E1A030BDF6 CRC64;  
 Query Match 71.1%; Score 59; DB 1; Length 378;  
 Best Local Similarity 68.8%; Pred. No. 0.015;  
 Matches 11; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 SYTEVTEGETEATAT 16  
 Db 321 AFVEVTEGEAAAT 336  
 RESULT 11  
 ID SCC1\_HUMAN STANDARD; PRT; 390 AA.  
 AC P29508; Q96J21;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Squamous cell carcinoma antigen 1 (SCCA-1) (Protein T4-A).  
 GN SERPINB3 OR SCCA1 OR SCCA.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE=92068241; PubMed=1958219;  
 RA Sumitani Y., Kishi F., Sekiguchi K., Kato H.;  
 RT "Squamous cell carcinoma antigen is a new member of the serine  
 RT protease inhibitors.";  
 RL Biochem. Biophys. Res. Commun. 181:51-58(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95241462; PubMed=7724531;  
 RA Schneider S.S., Schick C., Fish K.E., Miller E., Pena J.C.;  
 RA Treter S.D., Hui S.M., Silverman G.A.;  
 RT "A serine proteinase inhibitor locus at 18q21.3 contains a tandem  
 RT duplication of the human squamous cell carcinoma antigen gene.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:3147-3151(1995).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RL TISSUE=Lung;  
 RA Strausberg R.;  
 RL Submitted (Mar-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: MAY ACT AS A PROTEASE INHIBITOR TO MODULATE THE HOST  
 CC IMMUNE RESPONSE AGAINST TUMOR CELLS.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- TISSUE SPECIFICITY: SQUAMOUS CELLS.  
 CC -1- DEVELOPMENTAL STAGE: ITS EXPRESSION IS CLOSELY RELATED TO CELLULAR

CC DIFFERENTIATION IN BOTH NORMAL AND MALIGNANT SQUAMOUS CELLS.  
 CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: S66896; AAB20405.1; -  
 DR EMBL: U19556; AAA97552.1; -  
 DR EMBL: U19568; AAA86317.1; -  
 DR EMBL: U19559; AAA86317.1; JOINED.  
 DR EMBL: U19560; AAA86317.1; JOINED.  
 DR EMBL: U19562; AAA86317.1; JOINED.  
 DR EMBL: U19565; AAA86317.1; JOINED.  
 DR EMBL: U19567; AAA86317.1; JOINED.  
 DR EMBL: U19569; AAA86316.1; -  
 DR EMBL: U19559; AAA86316.1; JOINED.  
 DR EMBL: U19560; AAA86316.1; JOINED.  
 DR EMBL: BC005224; AAH05224.1; -  
 DR PIR: JTO966; JTO966.  
 DR HSSP: P01008; IATF.  
 DR GeneW: HGNC:10569; SERPINB3.  
 DR InterPro: IPR000215; Serpin.  
 DR Pfam: PF00079; serpin; 1.  
 DR SMART: SM00093; SERPIN; 1.  
 DR PROSITE: PS00284; SERPIN; 1.  
 KW Serpin; Serine protease inhibitor.  
 FT ACT\_SITE 354 355  
 FT CONFLICT 357 357  
 SQ SEQUENCE 390 AA; 44564 MW; E5F27F986C752CFA CRC64;  
 Query Match 71.1%; Score 59; DB 1; Length 390;  
 Best Local Similarity 68.8%; Pred. No. 0.015;  
 Matches 11; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 SYTEVTEGETEATAT 16  
 Db 332 AFVEVTEGEAAAT 347  
 RESULT 12  
 ID SCC2\_HUMAN STANDARD; PRT; 390 AA.  
 AC P48594;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Squamous cell carcinoma antigen 2 (SCCA-2) (Lepin).  
 GN SERPINB4 OR SCCA2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95241462; PubMed=7724531;  
 RA Schneider S.S., Schick C., Fish K.E., Miller E., Pena J.C.;  
 RA Treter S.D., Hui S.M., Silverman G.A.;  
 RT "A serine proteinase inhibitor locus at 18q21.3 contains a tandem  
 RT duplication of the human squamous cell carcinoma antigen gene.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:3147-3151(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96013887; PubMed=7589435;  
 RA Barnes R.C., Morrell D.M.;  
 RT "Identification of a novel human serpin gene: cloning sequencing and  
 RT expression of Lepin.";  
 RL FEBS Lett. 373:61-65(1995).

[3]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE-21167379; PubMed-11267667;  
 RA Hamada K., Shinomiya H., Asano Y., Kihana T., Iwamoto M., Hanakawa Y.,  
 RA Hashimoto K., Hirose S., Kyo S., Ito M.;  
 RT "Molecular cloning of human squamous cell carcinoma antigen 1 gene and  
 RT characterization of its promoter.";  
 RL Blochlm. Biophys. Acta 1518:124-131(2001).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Bone marrow;  
 RA Strauberg R.;  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: MAY ACT AS A PROTEASE INHIBITOR TO MODULATE THE HOST  
 CC IMMUNE RESPONSE AGAINST TUMOR CELLS.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- TISSUE SPECIFICITY: SQUAMOUS CELLS.  
 CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U19576; AAA92602.1; -;  
 DR EMBL: U19570; AAA92602.1; JOINED.  
 DR EMBL: U19571; AAA92602.1; JOINED.  
 DR EMBL: U19572; AAA92602.1; JOINED.  
 DR EMBL: U19574; AAA92602.1; JOINED.  
 DR EMBL: U19575; AAA92602.1; JOINED.  
 DR EMBL: U19557; AAA97553.1; -;  
 DR EMBL: X89015; CAA61420.1; -;  
 DR EMBL: AB035089; BAB21525.1; -;  
 DR EMBL: BC017401; AAL17401.1; -;  
 DR HSSP: P01008; IATH.  
 DR Genev. HGNC:10570; SERPINB4.  
 DR MIM: 600518; -;  
 DR InterPro: IPR000215; Serpin.  
 DR Pfam: PF00079; serpin; 1.  
 DR SMART: SM00093; SERPIN; 1.  
 DR PROSITE: PS00284; SERPIN; 1.  
 KW Serpin: Serine protease inhibitor.  
 FT ACT\_SITE 354 355  
 FT SEQUENCE 390 AA; 44854 MW; 04E213CD89258705 CRC64;  
 SO  
 Query Match 71.1%; Score 59; DB 1; Length 390;  
 Best Local Similarity 68.8%; Pred. NO. 0.015;  
 Matches 11; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
 Oy 1 SYIEVTEGTEATAT 16  
 Db 332 AFVEVTEGVEAAAT 347  
 RESULT 13  
 SPB8\_HUMAN STANDARD: PRT; 374 AA.  
 AC P50452;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Cytoplasmic antiprotease 2 (CAP2) (CAP-2) (Protease inhibitor 8)  
 DE (Serpin B9).  
 GN SERPINB8 OR P18.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OC NCBI\_TaxID=9606;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RP

RC TISSUE=Placenta;  
 RX MEDLINE-96102039; PubMed-8530382;  
 RA Sprecher C.A., Morgenstern K.A., Mathewes S., Dahlen J.R.,  
 RA Schrader S.K., Foster D.C., Kistiel W.;  
 RT "Molecular cloning, expression, and partial characterization of two  
 RT novel members of the ovalbumin family of serine proteinase  
 RT inhibitors.";  
 RL J. Biol. Chem. 270:29854-29861(1995).  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RA MEDLINE-97066975; PubMed-8910377;  
 RX Sun J., Bird C.H., Sutton V., McDonald L., Coughlin P.B., Jong T.A.,  
 RA Traipant J.A., Bird P.I.;  
 RT "A cytosolic granzyme B inhibitor related to the viral apoptotic  
 RT regulator cytokine response modifier A is present in cytotoxic  
 RT lymphocytes.";  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: LA0377; AAC41939.1; -;  
 DR HSSP: P05120; 1BY7  
 DR Genev. HGNC:8952; SERPINB8.  
 DR MIM: 601697; -;  
 DR InterPro: IPR000215; Serpin.  
 DR Pfam: PF00079; serpin; 1.  
 DR SMART: SM00093; SERPIN; 1.  
 DR PROSITE: PS00284; SERPIN; 1.  
 KW Serpin: Serine protease inhibitor.  
 FT ACT\_SITE 339 340  
 FT SEQUENCE 374 AA; 42786 MW; E855A033519AC60E CRC64;  
 SO  
 Query Match 69.9%; Score 58; DB 1; Length 374;  
 Best Local Similarity 73.3%; Pred. NO. 0.022;  
 Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 Oy 2 YIEVTEGTEATAT 16  
 Db 318 FVEVNEGTEAAAT 332  
 RESULT 14  
 SPB9\_HUMAN STANDARD: PRT; 376 AA.  
 AC P50453;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Cytoplasmic antiprotease 3 (CAP3) (CAP-3) (Protease inhibitor 9)  
 DE (Serpin B9).  
 GN SERPINB9 OR P19.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OC NCBI\_TaxID=9606;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RA MEDLINE-96102039; PubMed-8530382;  
 RX Sprecher C.A., Morgenstern K.A., Mathewes S., Dahlen J.R.,  
 RA Schrader S.K., Foster D.C., Kistiel W.;  
 RT "Molecular cloning, expression, and partial characterization of two  
 RT novel members of the ovalbumin family of serine proteinase  
 RT inhibitors.";  
 RL J. Biol. Chem. 270:29854-29861(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta; and Bone marrow;  
 RA MEDLINE-97066975; PubMed-8910377;  
 RX Sun J., Bird C.H., Sutton V., McDonald L., Coughlin P.B., Jong T.A.,  
 RA Traipant J.A., Bird P.I.;  
 RT "A cytosolic granzyme B inhibitor related to the viral apoptotic  
 RT regulator cytokine response modifier A is present in cytotoxic  
 RT lymphocytes.";

RL J. Biol. Chem. 271:27802-27809(1996).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Placenta;  
 RA Strausberg R.;  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: GRANZYME B INHIBITOR.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: LA0378; AAC41940.1; -  
 DR EMBL: U71364; AAC50793.1; -  
 DR EMBL: BC002538; AAC02538.1; -  
 DR HSSP: P05120; 1BY7.  
 DR Genew: HGNC:8955; SERPINB9.  
 DR MIM: 601799; -  
 DR InterPro: IPR000215; Serpin.  
 DR Pfam: PF00079; serpin.1.  
 DR SMART: SM00093; SERPIN; 1.  
 DR PROSITE: PS00284; SERPIN; 1.  
 KW Serpin; Serine protease inhibitor.  
 FT ACT\_SITE 340 341 REACTIVE BOND (BY SIMILARITY).  
 SQ SEQUENCE 376 AA; 42403 MW; ECA80869B939753 CRC64;  
 Query Match 69.9%; Score 58; DB 1; Length 376;  
 Best Local Similarity 68.8%; Pred. No. 0.022;  
 Matches 11; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
 Oy 1 SIEVEEGESTEAAT 16  
 Db 318 SFEVNEEGTEAAAS 333  
 RESULT 15  
 NEUS\_RAT  
 ID NEUS\_RAT STANDARD; PRT; 410 AA.  
 AC Q9JLD2; Q9JLD1;  
 DT 15-JUN-2002 (Rel. 41; Created)  
 DT 15-JUN-2002 (Rel. 41; Last sequence update)  
 DT 15-JUN-2002 (Rel. 41; Last annotation update)  
 DE Neuroserpin precursor (Protease inhibitor 17).  
 GN SERPIN11 OR P112.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Pituitary;  
 RX MEDLINE=20115094; PubMed=10642518;  
 RA Hill R.M.; Parmar P.K.; Coates L.C.; Mezey E.; Pearson J.F.,  
 RA Birch N.P.;  
 RT "Neuroserpin is expressed in the pituitary and adrenal glands and  
 RT induces the extension of neurite-like processes in Atf-20 cells.";  
 RL Blochem. J. 345:595-601(2000).  
 CC -1- FUNCTION: SERINE PROTEASE INHIBITOR THAT INHIBITS PLASMINOGEN  
 CC ACTIVATORS AND PLASMIN BUT NOT THROMBIN. MAY BE INVOLVED IN THE  
 CC FORMATION OR REORGANIZATION OF SYNAPTIC CONNECTIONS AS WELL AS FOR  
 CC SYNAPTIC PLASTICITY IN THE ADULT NERVOUS SYSTEM. MAY PROTECT  
 CC NEURONS FROM CELL DAMAGE BY TISSUE-TYPE PLASMINOGEN ACTIVATOR (BY  
 CC similarity).  
 CC -1- SUBCELLULAR LOCATION: Secreted (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: AF193014; AAF70386.1; -  
 DR EMBL: AF193015; AAF70387.1; -  
 DR HSSP: P05120; 1BY7.  
 DR InterPro: IPR000215; Serpin.  
 DR Pfam: PF00079; serpin.1.  
 DR SMART: SM00093; SERPIN; 1.  
 DR PROSITE: PS00284; SERPIN; 1.  
 KW Serpin; Serine protease inhibitor; Glycoprotein; Signal.  
 FT SIGNAL 1 16  
 FT CHAIN 17 410  
 FT ACT\_SITE 362 363 REACTIVE BOND (BY SIMILARITY).  
 FT CARBOHYD 157 157 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 321 321 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 401 401 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 12 12 L -> M (IN REF. 1; AAF70387).  
 FT CONFLICT 177 177 H -> Q (IN REF. 1; AAF70387).  
 SQ SEQUENCE 410 AA; 46278 MW; 99008624B3A36BD8 CRC64;  
 Query Match 68.7%; Score 57; DB 1; Length 410;  
 Best Local Similarity 64.7%; Pred. No. 0.035;  
 Matches 11; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 Oy 1 SIEVEEGESTEAATG 17  
 Db 340 SFEVNEEGSEAAVAGS 356

Search completed: July 11, 2003, 11:55:15  
 Job time : 2.62578 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: July 11, 2003, 11:44:31 ; Search time 7.42204 Seconds  
(Without alignments)  
471.946 Million cell updates/sec

Title: US-10-091-442-4  
Perfect score: 83  
Sequence: 1 SYIEVTEEGTEATATAC 17

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

## Database :

SPTREMBL\_21:\*  
1: sp.\_archaea:\*  
2: sp.\_bacteria:\*  
3: sp.\_fungi:\*  
4: sp.\_human:\*  
5: sp.\_invertebrate:\*  
6: sp.\_mammal:\*  
7: sp.\_mhc:\*  
8: sp.\_organelle:\*  
9: sp.\_phage:\*  
10: sp.\_plant:\*  
11: sp.\_rodent:\*  
12: sp.\_virus:\*  
13: sp.\_vertebrate:\*  
14: sp.\_unclassified:\*  
15: sp.\_virus:\*  
16: sp.\_bacteriap:\*  
17: sp.\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	72	86.7	334	12	Q9DHG4 yaba-like d
2	69	83.1	380	11	Q9D695 mus musculu
3	69	83.1	388	11	Q9CQV3 mus musculu
4	69	83.1	388	11	Q91212 mus musculu
5	68	81.9	379	11	Q9D7S8 mus musculu
6	68	81.9	379	11	Q9D1S4 mus musculu
7	68	81.9	387	11	Q9D105 mus musculu
8	68	81.9	388	11	Q922G2 mus musculu
9	66	79.5	379	5	Q9NFT6 anopheles g
10	66	79.5	379	5	Q8WSY0 anopheles g
11	66	79.5	380	5	Q9NFT5 anopheles g
12	66	79.5	380	5	Q8WSX8 anopheles g
13	66	79.5	392	5	Q917G5 drosophila
14	66	79.5	392	5	Q917G5 drosophila
15	66	79.5	418	5	Q8T0M5 drosophila
16	66	79.5	424	5	Q917G6 drosophila

17	66	79.5	431	5	Q9V991 drosophila
18	65	78.3	380	11	Q920J5 ratu
19	62	74.7	177	4	Q9BSM3 homo sapien
20	62	74.7	280	10	Q9XET7 avena fatua
21	62	74.7	374	5	Q9V3L3 drosophila
22	62	74.7	377	11	Q08804 mus musculu
23	62	74.7	382	11	Q8VHP7 mus musculu
24	62	74.7	411	10	Q94DW6 oryza sativ
25	61	73.5	374	5	Q9NGS0 bombyx mori
26	60	72.3	359	11	Q9DGA7 mus musculu
27	60	72.3	382	5	Q8WSX9 anopheles g
28	60	72.3	395	5	Q8WSX7 anopheles g
29	60	72.3	400	10	Q40076 hordeum vul
30	59	71.1	338	4	Q9BYR8 homo sapien
31	59	71.1	369	4	Q9BYR7 homo sapien
32	59	71.1	377	5	Q9GP13 Ixodes ricin
33	59	71.1	378	5	Q8WDX1 rhipicephal
34	59	71.1	379	11	Q9D0S8 mus musculu
35	59	71.1	385	11	Q8VHQ0 mus musculu
36	58	69.9	374	11	Q08800 mus musculu
37	58	69.9	380	5	Q8WDX0 rhipicephal
38	58	69.9	386	11	Q9D1E7 mus musculu
39	58	69.9	397	10	Q43492 hordeum vul
40	57	68.7	225	6	Q9GMA9 sus scrofa
41	57	68.7	373	5	Q9NH65 trichinella
42	57	68.7	374	16	Q8VYR8 anabaena sp
43	57	68.7	398	10	Q40066 hordeum vul
44	57	68.7	415	6	Q9GMA6 sus scrofa
45	57	68.7	459	11	Q9WTT1 cavia porce

## ALIGNMENTS

RESULT 1  
ID Q9DHG4 PRELIMINARY; PRT; 334 AA.  
AC Q9DHG4  
DT 01-MAR-2001 (TREMURel. 16, Created)  
DT 01-MAR-2001 (TREMURel. 16, Last sequence update)  
DT 01-JUN-2002 (TREMURel. 21, Last annotation update)  
DE 149R protein.  
GN 149R.  
OS yaba-like disease virus (YLDV).  
OC viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
OC yatapoxvirus.  
OC NCBI\_TaxID=132475;  
CC [1]  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Lee H.J.;  
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-21176366; PubMed-11277691;  
RA Lee H.J.; Essani K.; Smith G.L.;  
RT "The genome sequence of yaba-like disease virus, a yatapoxvirus.";  
RL Virology 281:170-192(2001).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Lee H.J.;  
RL Thesis (2000), Sir William Dunn School of Pathology, University of.  
CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.  
DR EMBL: AJ293568; CAC21387.1; .  
DR HSSP; P05120; 18Y7.  
DR InterPro: IPR00215; Serpin.  
DR Pfam: PF00079; serpin; 1.  
DR SMART: SMO0093; SERPIN; 1.  
DR PROSITE: PS00284; SERPIN; 1.  
KW Serpin.  
SQ SEQUENCE 334 AA; 38190 MW; D7FD004191B3C19F CRC64;  
Query Match 86.7%; Score 72; DB 12; Length 334;  
Best Local Similarity 93.8%; Pred. No. 0.00075;

```

Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 SYIEVTEGTEATAAT 16
Db 277 SYIEVTEGTEATAAT 292

RESULT 2
ID 09D695 PRELIMINARY; PRT: 380 AA.
AC 09D695;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 4631416M05R1K protein (megsin).
GN SERPINB7 OR 4631416M05R1K.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=SKIN;
RA MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Oikido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima M., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Wetz C., Whittaker C., Wilming L.,
RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21368006; PubMed=11473647;
RA Nangaku M., Miyata T., Suzuki D., Umezono T., Hashimoto T., Wada T.,
RA Yagi M., Nagano N., Inagi R., Kurokawa K.;
RT "Cloning of rodent mesgin revealed its up-regulation in
RT mesangioliproliferative nephritis.";
RL Kidney Int. 60:641-652(2001).
DR EMBL: AK014524; BAB29410.1; -
DR EMBL: AF105328; AAL16768.1; -
DR HSSP: P05121; 1A7C.
DR MGD: MGI:2151053; Serpinb7.
DR InterPro: IPR000215; Serpin.
DR Pfam: PF00079; serpin.1.
DR SMART: SM00093; SERPIN.1.
DR PROSITE: PS00284; SERPIN; UNKNOWN.1.
SQ SEQUENCE 380 AA; 43050 MW; C9240272BCFB9C64 CRC64;

Query Match 83.1%; Score 69; DB 11; Length 380;
Best Local Similarity 87.5%; Pred. No. 0.0027;
Matches 14; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 SYIEVTEGTEATAAT 16
Db 325 SYIEVTEGTEATAAT 340

```

RESULT 3  
09C0V3

```

ID 09C0V3 PRELIMINARY; PRT: 388 AA.
AC 09C0V3;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 2310046M08R1K protein.
GN 2310046M08R1K.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TONGUE;
RA MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Oikido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Lyons P., Marchionni L., Mashima M., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Wetz C., Whittaker C., Wilming L.,
RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK009855; BAB26546.1; -
DR EMBL: AK009003; BAB26017.1; -
DR HSSP: P05619; 1HLE.
DR MGD: MGI:1914207; 2310046M08R1K.
DR InterPro: IPR000215; Serpin.
DR Pfam: PF00079; serpin.1.
DR SMART: SM00093; SERPIN.1.
SQ SEQUENCE 388 AA; 43482 MW; 533D8282F58F09AE CRC64;

Query Match 83.1%; Score 69; DB 11; Length 388;
Best Local Similarity 76.5%; Pred. No. 0.0028;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 1 SYIEVTEGTEATAATG 17
Db 332 SYIVNEGTEATAATG 348

RESULT 4
ID 091212 PRELIMINARY; PRT: 388 AA.
AC 091212;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 2310046M08R1K gene.
GN 2310046M08R1K.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC Submitted R.;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -i SIMILARITY: BELONGS TO THE SERPIN FAMILY.
DR EMBL: BC010313; AAH10313.1; -
DR MGD: MGI:1914207; 2310046M08R1K.

```

DR InterPro: IPR000215; Serpin.  
DR Pfam: PF00079; serpin. 1.  
KW Serpin.  
SQ SEQUENCE 388 AA; 43544 MW; AF4BB62FB084E6FA CRC64;

Query Match 83.1%; Score 69; DB 11; Length 388;  
Best Local Similarity 76.5%; Pred. No. 0.0028;  
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 SYIEVTEEGTEATATG 17  
DB 332 SFVEVNEEGTEAAATG 348

RESULT 5  
O9D7S8 PRELIMINARY; PRT; 379 AA.  
AC O9D7S8;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE 1190005M04R1k protein.  
GN SERPINB1A OR 1190005M04R1K.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;

RA [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6J; TISSUE-STOMACH;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Atzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
RA Schirml L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Welter C., Whitaker C., Wilming L.,  
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
RA Hayashizaki Y.,  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.  
DR EMBL: AK008914; BAB25964.1; -;  
DR HSSP: P05619; 1HLE.  
DR MGD: MGI:1913472; Serpinb1a.  
DR InterPro: IPR000215; Serpin.  
DR Pfam: PF00079; serpin. 1.  
DR SMART: SM00093; SERPIN; 1.  
DR PROSITE: PS00284; SERPIN; 1.  
KW Serpin.  
SQ SEQUENCE 379 AA; 42580 MW; 193728585077AB4B CRC64;

Query Match 81.9%; Score 68; DB 11; Length 379;  
Best Local Similarity 76.5%; Pred. No. 0.004;  
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 SYIEVTEEGTEATATG 17  
DB 322 SFVEVNEEGTEAAATG 338

RESULT 6  
O9D154

ID O9D154 PRELIMINARY; PRT; 379 AA.  
AC O9D154;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE 1190005M04R1k protein (RIKEN cDNA 1190005M04 gene) (E1A).  
GN SERPINB1A OR 1190005M04R1K OR SERPINB1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;

RA [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6J; TISSUE-EMBRYO;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Atzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
RA Schirml L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Kamiya M., Lee N.H.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Welter C., Whitaker C., Wilming L.,  
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
RA Hayashizaki Y.,  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.  
DR EMBL: AK003930; BAB23079.1; -;  
DR HSSP: P05619; 1HLE.  
DR MGD: MGI:1913472; Serpinb1a.  
DR InterPro: IPR000215; Serpin.  
DR Pfam: PF00079; serpin. 1.  
DR SMART: SM00093; SERPIN; 1.  
DR PROSITE: PS00284; SERPIN; 1.  
KW Serpin.  
SQ SEQUENCE 379 AA; 42575 MW; F492F1EEB88E80B1 CRC64;

Query Match 81.9%; Score 68; DB 11; Length 379;  
Best Local Similarity 76.5%; Pred. No. 0.004;  
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 SYIEVTEEGTEATATG 17  
DB 322 SFVEVNEEGTEAAATG 338

RESULT 7  
O9D105

```

ID 09D105 PRELIMINARY; PRT: 387 AA.
AC 09D105;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE 18 days embryo cDNA, RIKEN full-length enriched library,
DE clone:1110001H02, full insert sequence.
GN SCCA2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Glass C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirni L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boileau D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Norone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Wetz C., Whitaker C., Wilming L.,
RA Wysshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.
DR EMBL: AK003220; BAB22650.1; -.
DR HSSP: P05120; 1B77.
DR MGD: MGI:127952; Scca2.
DR InterPro: IPR000215; Serpin.
DR Pfam: PF00079; serpin. 1.
DR SMART: SM00093; SERPIN; 1.
DR PROSITE: PS00284; SERPIN; 1.
KW Serpin.
SQ SEQUENCE 387 AA; 44558 MW; AE25C02520F3B194 CRC64;

Query Match 81.9%; Score 68; DB 11; Length 387;
Best Local Similarity 76.5%; Pred. No. 0.0041;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 SYIEVTEGTEATATG 17
DB 330 SFVEVNEGTEAATG 346

```

```

RA Silverman G.A.;
RT "A murine ortholog of the human serpin SCCA2 maps to chromosome 1 and
RT inhibits chymotrypsin-like serine proteinases.";
RL Genomics 54:297-306(1998).
CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.
DR EMBL: AF063937; AAC95432.1; -.
DR HSSP: P05120; 1B77.
DR MGD: MGI:127952; Scca2.
DR InterPro: IPR000215; Serpin.
DR Pfam: PF00079; serpin. 1.
DR SMART: SM00093; SERPIN; 1.
DR PROSITE: PS00284; SERPIN; 1.
KW Serpin.
SQ SEQUENCE 388 AA; 44573 MW; D77AFA1223EB9CD CRC64;

Query Match 81.9%; Score 68; DB 11; Length 388;
Best Local Similarity 76.5%; Pred. No. 0.0041;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 SYIEVTEGTEATATG 17
DB 331 SFVEVNEGTEAATG 347

RESULT 9
O9NFT6 PRELIMINARY; PRT: 379 AA.
AC 09NFT6;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Putative serine protease inhibitor.
GN SP11A.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
OC Anopheles.
OX NCBI_TaxID=7165;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=4A R/R;
RA Danielel A., Kafatos F.C.;
RT "Serine protease inhibitors of Anopheles gambiae.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ271352; CAB69784.1; -.
DR HSSP: P01008; 1A7H.
DR InterPro: IPR000215; Serpin.
DR Pfam: PF00079; serpin. 1.
DR SMART: SM00093; SERPIN; 1.
DR PROSITE: PS00284; SERPIN; UNKNOWN_1.
KW Protease.
SQ SEQUENCE 379 AA; 42157 MW; 3DE2DFOA07BCB959 CRC64;

Query Match 79.5%; Score 66; DB 5; Length 379;
Best Local Similarity 76.5%; Pred. No. 0.0087;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 SYIEVTEGTEATATG 17
DB 323 AFIEVNEGTEAATG 339

RESULT 10
O8WSY0 PRELIMINARY; PRT: 379 AA.
AC 08WSY0;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Serpin.
GN SP121F.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

```





RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Morley K.C., Mu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*."  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003790; AAC22209.1; -;  
 DR HSSP: P01008; IATH.  
 DR FLYBASE: FBgn0028985; sp4.  
 DR InterPro: IPR000886; ER\_target.  
 DR InterPro: IPR000215; Serpin.  
 DR Pfam: PF00079; serpin.1.  
 DR SMART: SM00093; SERPIN.1.  
 DR PROSITE: PS00014; ER\_TARGET; UNKNOWN\_1.  
 SQ SEQUENCE 392 AA; 44127 MW; 5D221529C5E0822E CRC64;

Query Match 79.5%; Score 66; DB 5; Length 392;  
 Best Local Similarity 76.5%; Pred. No. 0.009;

Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 SYIEVTEGTEATATG 17  
 DB 319 AFIEVNEGTETAAATG 335

## RESULT 14

090115

ID 090115 PRELIMINARY; PRT; 392 AA.

AC 090115; 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE Serine protease inhibitor (Serpin-4).

GN SP4 OR CG9453.

OS *Drosophila melanogaster* (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RA Han J.H., Zhang H., Min G.S., Hashimoto C.;

RT "Identification of functional *Drosophila* serpins."

RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.

DR EMBL: AJ251747; CAB6309.1; -;

DR HSSP: P01008; IATH.

DR FLYBASE: FBgn0028985; sp4.

DR InterPro: IPR000886; ER\_target.

DR InterPro: IPR000215; Serpin.

DR Pfam: PF00079; serpin.1.

DR SMART: SM00093; SERPIN.1.

DR PROSITE: PS00014; ER\_TARGET; UNKNOWN\_1.

KW Protease; Serpin.

SQ SEQUENCE 392 AA; 44133 MW; 0A01CF14D3B85B9F CRC64;

Query Match 79.5%; Score 66; DB 5; Length 392;  
 Best Local Similarity 76.5%; Pred. No. 0.009;

Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 SYIEVTEGTEATATG 17  
 DB 319 AFIEVNEGTETAAATG 335

## RESULT 15

08T0M5

ID 08T0M5 PRELIMINARY; PRT; 418 AA.

AC 08T0M5;

DT 01-JUN-2002 (TREMBlrel. 21, Created)

DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE GH21896p.  
 GN SP4.  
 OS *Drosophila melanogaster* (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RA Stapleton M., Broksstein P., Hong L., Agbayani A., Carlson J.,  
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
 RA Nuncio J., Paclob J., Paragas V., Park S., Phouanavong S., Wan K.,  
 RA Yu C., Lewis S.E., Rubin G.M., Celisner S.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AY069173; AAL39318.1; -;  
 SQ SEQUENCE 418 AA; 46961 MW; D956A2C04339B5E0 CRC64;

Query Match 79.5%; Score 66; DB 5; Length 418;  
 Best Local Similarity 76.5%; Pred. No. 0.0096;

Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 SYIEVTEGTEATATG 17  
 DB 358 AFIEVNEGTETAAATG 374

Search completed: July 11, 2003, 11:58:57  
 Job time : 10.422 secs

GenCore version 5.1.6  
Copyright (c) 1993  
Combining  
16424 Second  
Agments)  
million cell upda  
Copyright using sw model  
42:15

908470

Scoring tab  
atch 1008  
first 45 summaries

## SUMMARIES

Description

## ALIGNMENTS

lettuce resistance  
lettuce pest resis  
drosophila melanog  
C albicans apertosis  
Helicobacter pylori  
Helicobacter aerugi  
pseudomonas endonii  
P. furiosus human prot  
Novel human protein  
Human protein segm  
Novel human proteol  
Prophila abacterium  
Chlamydia pneumonia  
Human cAMPX protein  
pyrodictum occul  
pyrodictum melanog  
Drosophila secreted pro  
Human  
Aspergillus niger  
Aspergillus melanog  
Drosophila melanog  
Wild type E. blati  
E. blati strain 1,3-pa  
E. blati fastidiosu  
Xylella polyphloede  
Human protein segm  
Human CBP (clone)  
Rice virus DNA poly  
P. abuyasi DNA po  
P. octatum DNA po  
P. octatum occul  
Pyron myosin Ila  
Human myosin Ila  
Novel human diapo  
Dipeptide transpo  
H. pylori ORF 024P  
H. pylori ORF  
Human protein segm  
Arabidopsis thaliana

RESULT 1	
AA848379	standard; Protein; 380 AA.
ID	AA848379
XX	PARA8379:
XX	16-AUG-1994 (first entry)
XX	Human megakaryocyte differentiation factor.
XX	MDF; thrombopoietin.
XX	Human megakaryocyte differentiation factor; platelet;
XX	Human megakaryocytic stimulating factor; cancer chemotherapy.
XX	Human megakaryocytic stimulating factor; cancer chemotherapy.
KW	haematopoietic stimulation factor; cancer chemotherapy.
KM	bone marrow transplantation; cancer chemotherapy.
KM	bone marrow transplantation; cancer chemotherapy.
OS	Homo sapiens..
XX	EP583884-A.
XX	23-FEB-1994.
PN	93EP-0305654.
PD	19-JUL-1993:
XX	92JP-0212305.
XX	17-JUL-1992:
PF	93JP-0067339.
XX	04-MAR-1993:
PR	(SUNR) TSUJIMOTO M.
XX	(TSUJ/) TSUJIMOTO M.
PA	(SUNR) SUNTORY LTD.
XX	Iwasa F., Kurihara T., Miura K., Nakazato H;
XX	Tsuji moto M., Tsurioka N., Yamaguchi N., Yamachi K;
PI	Tsuji moto M., Tsurioka N., Yamaguchi N., Yamachi K;
XX	WI: 1994-058782/08.
XX	WI: 1994-058782/08.
DR	N-PSDB; AAQ5670.
DR	N-PSDB; AAQ5670.

us-10-091-442-5.1rag

Fri Jul 11 15:11:04 2003

XX New megakaryocyte differentiation factor - isolated from human  
PT epidermoid carcinoma cells, used to treat conditions involving a  
PT decrease in platelets  
XX Claim 7: Page 30-32; 47pp; English.

XX Human MDF can be isolated from a culture of human epidermoid  
XX carcinoma A431 cells in protein-free medium. The MDF stimulates  
CC differentiation of megakaryocytes from myeloid cells in the presence  
CC of IL-3. The MDF acts in vivo as a thrombopoietin making it useful  
CC for treatment of diseases involving a decrease in platelet number  
CC (esp. thrombocytopenia) such as occurs in bone marrow  
CC transplantation and in chemotherapy. MDF has mol.wt. 55-57kD by  
CC SDS-PAGE and exhibits an isoelectric point of 6.0-7.0.  
XX Sequence 380 AA: DB 15; Length 380; Gaps 0;  
XX Query Match Similarity 100.0%; Score 44; DB 15; Length 380;  
XX Best Local Similarity 100.0%; Pred. No. 1.1; Indels 0; Gaps 0;  
Matches 9; Conservative

OY 1 QYLRALGK 9  
DB 289 QYLRALGK 297

RESULT 2  
ID AAY08254 standard: Protein: 380 AA.  
XX AAY08254:  
XX AC  
XX 14-JUL-1999 (first entry)  
XX DT  
XX Human megalin protein.  
XX DE  
XX Megalin: mesangial cell; treatment; diagnosis; disease; Iga nephropathy;  
XX human; rat; murine.  
XX sapiens.

DR 9  
XX N-98WO-JP04269.  
XX PT  
XX Megalin pI: 8-0275302.  
XX PS  
XX Claim 1: Page 62  
XX CC  
XX CC This invention describes  
XX CC proteins from human, rat and  
XX CC specifically in mesangial cells  
XX CC useful for the treatment and diagnosis  
XX CC cells, such as Iga nephropathy.  
XX Sequence 380 AA:  
XX Query Match  
XX Best Local Similarity 100.0%; Score 44; DB 20; Length 380;  
Matches 9; Conservative 0; Mismatches 0; Indels 0;

OY 1 QYLRALGK 9  
DB 289 QYLRALGK 297

RESULT 3  
ID AAB24142 standard: Protein: 380 AA.  
XX AAB24142:  
XX AC  
XX 30-JAN-2001 (first entry)  
XX DT  
XX Human megalin protein sequence SEQ ID NO:2.  
XX DE  
XX Megalin: mesangium-predominant gene; serpin regulated; nephropathy;  
XX Iga; immunoglobulin A; detection; renal function; renal disorder;  
XX diagnosis; biological sample; blood; urine.  
XX Homo sapiens.  
XX OS  
XX WC200057189-A1.  
XX PN  
XX 28-SEP-2000.  
XX PD  
XX 17-MAR-2000; 2000WO-JP01646.  
XX PE  
XX 19-MAR-1999; 99JP-0075305.  
XX PR  
XX 28-OCT-1999; 99JP-0306623.  
XX PA  
XX (KURO/) KUROKAWA K.  
XX PA (FUSO) FUSO PHARM IND LTD.  
XX PA (MIYA/) MIYATA T.  
XX MIYATA T.  
XX PI  
XX WPI: 2000-611642/58.  
XX DR N-PSB; AAB99294.  
XX DT  
XX

XX Example 2: Page 66-69; 93pp; Japanese.  
XX PT  
XX biological sample  
XX DT  
XX Evaluating renal function comprises assaying megalin protein in  
XX biological sample  
XX PS  
XX The present invention describes a method for evaluating renal function.  
XX CC The method comprises assaying megalin protein in biological sample. Also  
XX CC described are: (1) use of a anti-megalin protein antibody for diagnosing;  
XX CC renal function; and (2) a kit for detecting megalin protein particles;  
XX CC (a) anti-megalin protein antibody attached to solid magnetic particles; and  
XX CC (b) direct or indirect fixing for the antibody to the particles;  
XX CC (c) a magnet. The process is useful for evaluating renal function and  
XX CC diagnosing renal disorders by assaying megalin protein in biological  
XX CC samples (preferably urine or blood). The process represents the human megalin  
XX CC protein, which is given in the exemplification of the present invention  
XX CC  
XX Sequence 380 AA:  
XX Query Match  
XX Best Local Similarity 100.0%; Score 44; DB 21; Length 380;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

RESULT 4  
ID AAB83075 standard: Protein: 380 AA.  
XX AAB83075:  
XX AC

```

XX 10-JUL-2001 (first entry)
DT
XX Human mesgin protein.
DE
XX Human; mesgin; mesangial cell proliferative nephritis; nephrotropic;
KW transgenic mouse; glomerular disease; animal model; drug screening.
XX
OS Homo sapiens.
XX
PN WO200124628-A1.
PD
XX 12-APR-2001.
PF
XX 06-OCT-2000: 2000WO-JP06988.
PR
XX 06-OCT-1999: 99JP-0285736.
PA (KURO/) KUROKAWA K.
PA (MIYA/) MIYATA T.
PI Miyata T.
PI
XX WPI: 2001-300136/31.
DR N-PSDB; AAF82438.
DR
XX
XX Mouse model for mesangial cell proliferative nephritis for development
PT and screening of new treatments -
XX
XX Example 4: Page 44-46; 62pp; Japanese.
PS
XX The present sequence is human mesgin. The human mesgin coding
CC sequence may be introduced into a mouse to produce an animal model of
CC mesangial cell proliferative nephritis. The symptoms include
CC enlargement of the mesangial base region, sedimentation of an immune
CC complex and an increase in mesangial cells. The animal model is useful
CC for analysing the pathology of chronic glomerular diseases and for
CC screening compositions for prevention and treatment of the diseases.
CC Highly uniform models can be made easily and in large numbers using
CC this method.
XX
SQ Sequence 380 AA:
Query Match 100.0%; Score 44; DB 22; Length 380;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 QYRALGLK 9
DB 289 QYRALGLK 297
RESULT 5
AA57111
ID AAR57111 standard; peptide; 9 AA.
XX
AC AAR57111;
XX
DT 16-AUG-1994 (first entry)
XX
DE Human megakaryocyte differentiation factor peptide 5.
XX
KW Human megakaryocyte differentiation factor; MDF; thrombopoietin;
KW haematopoietic stimulating factor; thrombocytopoena; platelet;
KW bone marrow transplantation; cancer chemotherapy.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 1 /note= "not determined"
XX
PN EP583884-A.

```

```

XX 23-FEB-1994.
PD
XX 19-JUL-1993: 93EP-0305654.
PF
XX 17-JUL-1992: 92JP-0212305.
PR 04-MAR-1993: 93JP-0067339.
XX
PA (SUNR ) SUNTORY LTD.
PA (TSUJ/) TSUJIMOTO M.
PI
XX Ishida N, Iwasa F, Kurihara T, Miura K, Nakazato H;
PI Tsujimoto M, Tsuruoka N, Yamaguchi N, Yamachi K;
XX
DR WPI: 1994-058782/08.
DR
XX New megakaryocyte differentiation factor - isolated from human
PT epidermoid carcinoma cells, used to treat conditions involving a
PT decrease in platelets
XX
XX Claim 1; Page 19; 47pp; English.
PS
XX Human MDF (see AAR48379) can be isolated from a culture of human
CC epidermoid carcinoma A431 cells in protein-free medium. The MDF
CC stimulates differentiation of megakaryocytes from myeloid cells
CC in the presence of IL-3. The MDF acts in vivo as a thrombopoietin
CC making it useful for treatment of diseases involving a decrease
CC in platelet number (esp. thrombocytopenia) such as occurs in bone
CC marrow transplantation and in chemotherapy. MDF has mol.wt.
CC 55-57KD by SDS-PAGE, exhibits an isoelectric point of 6.0-7.0 and
CC contains an amino acid sequence comprising at least one of the
CC sequences AAR57107-R57115.
XX
SQ Sequence 9 AA:
Query Match 88.6%; Score 39; DB 15; Length 9;
Best Local Similarity 100.0%; Pred. No. 7.8e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 2 YLRALGLK 9
DB 2 YLRALGLK 9
RESULT 6
AA99661
ID AAY99661 standard; Protein; 694 AA.
XX
AC AAY99661;
XX
DT 03-NOV-2000 (first entry)
XX
DE Human GTPase associated protein-12.
XX
KW Guanine nucleotide binding protein; GTP-binding protein; G-protein;
KW GTPase; GTPase associated protein; GTPAP; cell proliferation;
KW autoimmune; inflammatory; immune system disorder; cancer; AIDS;
KW acquired immune deficiency syndrome; asthma; atherosclerosis;
KW arthritis; systemic lupus erythematosus; psoriasis; human.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 10..24 /label= Beta_transducin
FT Modified-site 57 /note= "Potential phosphorylation site"
FT Modified-site 67 /note= "Potential phosphorylation site"
FT Domain 96..110 /label= Beta_transducin
FT Modified-site 99 /note= "Potential phosphorylation site"
FT

```

FT Modified-site 150 /note= "Potential phosphorylation site"  
 FT Modified-site 156 /note= "Potential phosphorylation site"  
 FT Modified-site 209 /note= "Potential phosphorylation site"  
 FT Modified-site 285 /note= "Potential phosphorylation site"  
 FT Modified-site 331 /note= "Potential phosphorylation site"  
 FT Modified-site 346 /note= "Potential phosphorylation site"  
 FT Modified-site 360 /note= "Potential phosphorylation site"  
 FT Modified-site 388 /note= "Potential phosphorylation site"  
 FT Modified-site 416 /note= "Potential phosphorylation site"  
 FT Modified-site 430 /note= "Potential phosphorylation site"  
 FT Modified-site 467 /note= "Potential phosphorylation site"  
 FT Modified-site 477 /note= "Potential phosphorylation site"  
 FT Modified-site 500 /note= "Potential phosphorylation site"  
 FT Modified-site 522 /note= "Potential phosphorylation site"  
 FT Modified-site 550 /note= "Potential phosphorylation site"  
 FT Modified-site 584 /note= "Potential phosphorylation site"  
 FT Modified-site 688 /note= "Potential phosphorylation site"  
 FT Modified-site /note= "Potential phosphorylation site"  
 PN WO20031263-A2.  
 PD 02-JUN-2000.  
 XX 23-NOV-1999; 99WO-US28013.  
 XX 23-NOV-1998; 98US-0109592.  
 PR 04-FEB-1999; 99US-0118610.  
 PR 06-APR-1999; 99US-0127990.  
 XX (INCY-) INCYTE PHARM INC.  
 XX Hillman JL, Tang YT, Bandman O, Lal P, Yue H, Lu DM, Baughn MR;  
 PI Yang J, Azimzal Y;  
 DR WPI: 2000-400073/34.  
 DR N-PSDB; AAA49182.  
 XX Human GTPase associated proteins, polynucleotides, and antibodies,  
 PT useful for diagnosing, preventing and treating various diseases such as  
 PT atherosclerosis, cancer, acquired immune deficiency syndrome (AIDS),  
 PT asthma, and autoimmune diseases -  
 XX  
 PS Claim 1; Page 94-96; 144pp; English.  
 CC Human cDNA libraries from various tissues were screened for GTPase  
 CC associated proteins (GTPAP). The present sequence is human  
 CC GTPAP-12 protein. This sequence was derived from a cDNA library of the  
 CC prostate tumour tissue removed from a 65 year old male.  
 CC This protein is expressed in reproductive, haematopoietic/immune and  
 CC gastrointestinal tissue. The GTPAP proteins may be used to define  
 CC agonists and antagonists of GTPAP activity and to generate antibodies  
 CC to GTPAP. This means the GTPAP proteins may be useful for treatment or  
 CC prevention of diseases associated with GTPAP such as cell proliferation  
 CC disorders, autoimmune disorders, inflammatory disorders, immune system  
 CC disorders, cancer, AIDS, asthma, atherosclerosis, arthritis, systemic  
 CC lupus erythematosus and psoriasis.  
 CC

SQ Sequence 694 AA;  
 Query Match 79.5%; Score 35; DB 21; Length 694;  
 Best Local Similarity 87.5%; Pred. No. 1.3e+02;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QYLRALGL 8  
 :|||||||  
 DB 551 RYLRALGL 558  
 RESULT 7  
 ABG06229  
 ID ABG06229 standard; Protein; 213 AA.  
 AC ABG06229;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #6220.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 OS Homo sapiens.  
 PN WO200175067-A2.  
 PD 11-OCT-2001.  
 XX 30-MAR-2001; 2001WO-US08631.  
 PF 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX (HISE-) HISEQ INC.  
 XX Drmanac RT, Liu C, Tang YT;  
 PI  
 DR WPI: 2001-639362/73.  
 DR N-PSDB; AAS70416.  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 PS Claim 20; SEQ ID No 36588; 103pp; English.  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 SQ Sequence 213 AA;

Query Match 77.3%; Score 34; DB 22; Length 213;  
 Best Local Similarity 87.5%; Pred. No. 61;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 OYLRALGL 8  
 DB 142 OYLRALGL 149

## RESULT 8

AA06788  
 ID AAY06788 standard; Protein; 250 AA.

AC AAY06788;

DT 22-JUN-1999 (first entry)

DE C. amalonaticus maleate isomerase.

KM Maleate isomerase; recombinant; P. fluorescens; C. amalonaticus;  
 KW L-aspartic acid.

OS Citrobacter amalonaticus.

PN EP908520-A2.

PD 14-APR-1999.

PF 08-OCT-1998; 98EP-0308213.

PR 08-OCT-1997; 97JP-0276261.

PA (JAPC ) NIPPON SHOKUBAI CO LTD.

PI Komatsuzaki S, Mukoyama M, Yasuda S;

DR WPI; 1999-217071/19.

DR N-PSDB; AAX32470.

PT New maleate isomerase useful for production of L-aspartic acid

PS Claim 6; Page 11-12; 21pp; English.

XX This represents a C. amalonaticus maleate isomerase. Host cells

CC transformed with a vector comprising the DNA are used for the recombinant

CC expression of the protein. The maleate isomerases of the invention

CC isolated from P. fluorescens and C. amalonaticus are useful for

CC producing L-aspartic acid.

SO Sequence 250 AA;

OY 1 OYLRALGL 9  
 DB 228 OYLRALGL 236

## RESULT 9

AA064160  
 ID AAM64160 standard; Protein; 1365 AA.

AC AAM64160;

DT 19-JUL-1999 (first entry)

DE Lettuce resistance protein RLGA.

KM Resistance gene; RLGA; RG2A; lettuce; disease resistance;

KW pest resistance; virus; fungus; protozoan; bacterium; nematode;  
 crop protection; transgenic plant.

XX Lactuca sativa.

OS Lactuca sativa.

XX Key Location/Qualifiers

FT Misc-difference 413 /note= "encoded by RAC"

FT Misc-difference 434 /note= "encoded by NTG"

FT Misc-difference 1057 /note= "encoded by NAG"

FT Misc-difference 1086 /note= "encoded by NAG"

FT Misc-difference 1086 /note= "encoded by NAG"

FT Misc-difference 1086 /note= "encoded by NAG"

FT Misc-difference 1086 /note= "encoded by NAG"

FT Misc-difference 1086 /note= "encoded by NAG"

FT Misc-difference 1086 /note= "encoded by NAG"

FT Misc-difference 1086 /note= "encoded by NAG"

FT Misc-difference 1086 /note= "encoded by NAG"

FT Misc-difference 1086 /note= "encoded by NAG"

FT Misc-difference 1086 /note= "encoded by NAG"

FT Misc-difference 1086 /note= "encoded by NAG"

FT Misc-difference 1086 /note= "encoded by NAG"

FT Misc-difference 1086 /note= "encoded by NAG"

FT Misc-difference 1086 /note= "encoded by NAG"

FT Misc-difference 1086 /note= "encoded by NAG"

FT Misc-difference 1086 /note= "encoded by NAG"

FT Misc-difference 1086 /note= "encoded by NAG"

FT Misc-difference 1086 /note= "encoded by NAG"

FT Misc-difference 1086 /note= "encoded by NAG"

FT Misc-difference 1086 /note= "encoded by NAG"

FT Misc-difference 1086 /note= "encoded by NAG"

FT Misc-difference 1086 /note= "encoded by NAG"

FT Misc-difference 1086 /note= "encoded by NAG"

FT Misc-difference 1086 /note= "encoded by NAG"

FT Misc-difference 1086 /note= "encoded by NAG"

FT Misc-difference 1086 /note= "encoded by NAG"

FT Misc-difference 1086 /note= "encoded by NAG"

FT Misc-difference 1086 /note= "encoded by NAG"

FT Misc-difference 1086 /note= "encoded by NAG"

FT Misc-difference 1086 /note= "encoded by NAG"

FT Misc-difference 1086 /note= "encoded by NAG"

FT Misc-difference 1086 /note= "encoded by NAG"

```

XX  US6350933-B1.
XX  26-FEB-2002.
XX  09-JAN-1998; 98US-0004838.
XX  10-JAN-1997; 97US-0781734.
XX  (REGC ) UNIV CALIFORNIA.
XX  Michelmore RW, Shen KA, Meyers BC;
XX  WPI: 2002-314681/35.
XX  N-PSDB; ABR67796.
XX  Isolated nucleic acid molecule, RG2B (resistance gene 2B)
XX  polynucleotide, encodes polypeptide useful for conferring resistance to
XX  plant downy mildew.
XX  Disclosure: Column 85-92; 209pp; English.
XX  CC  The invention relates to an isolated nucleic acid molecule comprises an
XX  CC  RG2B (resistance gene 2B) polynucleotide encoding a polypeptide
XX  CC  conferring resistance to plant downy mildew, hybridises to a fully
XX  CC  defined sequence of 15062 base pairs as given in the specification, under
XX  CC  a wash with 0.2 x SSC (standard sodium citrate) at 65plusOC. Also
XX  CC  included is a transgenic plant comprising a heterologous expression
XX  CC  cassette containing a promoter operably linked to an RG2B polynucleotide
XX  CC  encoding a polypeptide conferring resistance to plant downy mildew and
XX  CC  hybridises to a fully defined sequence of 1323 amino acids as
XX  CC  given in the specification, under a wash with 0.2 x SSC at 65plusOC.
XX  CC  Also disclosed are other pest resistance genes and their protein
XX  CC  products. The isolated nucleic acid molecule is used to confer resistance
XX  CC  to downy mildew in a plant. This comprises introducing into the plant
XX  CC  an expression cassette comprising a promoter operably linked to an RG2B
XX  CC  polynucleotide. The present sequence is a lettuce resistance
XX  CC  protein sequence (or protein consensus sequence).
XX  SQ  Sequence 1366 AA;
XX  QY 1 QYLRALGK 9
XX  DB 31 QYRREMGK 39
XX  Query Match 75.0%; Score 33; DB 23; Length 1366;
XX  Best Local Similarity 55.6%; Pred. No. 6.5e+02;
XX  Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
XX  RESULT 11
XX  AAW64176
XX  ID AAW64176 standard; Protein: 1890 AA.
XX  AC AAW64176;
XX  DT 19-JUL-1999 (first entry)
XX  DE Lettuce resistance protein RG2A.
XX  KW Resistance gene; RG2A; lettuce; disease resistance;
XX  KW pest resistance; virus; fungus; protozoan; bacterium; nematode;
XX  KW crop protection; transgenic plant.
XX  OS Lactuca sativa.
XX  FT Key Location/Qualifiers
XX  FT Misc-difference 1497
XX  FT /note= "encoded by TAA"
XX  FT Misc-difference 1884
XX  FT /note= "encoded by TAA"
XX  PN W09830083-A1.

```

```

XX  16-JUL-1998.
XX  09-JAN-1998; 98WO-US000615.
XX  10-JAN-1997; 97US-0781734.
XX  (REGC ) UNIV CALIFORNIA.
XX  Meyers B, Michelmore RW, Shen K;
XX  WPI: 1998-398692/34.
XX  N-PSDB; AAV44231.
XX  New resistance gene nucleic acids - useful to produce disease
XX  PT resistant plants, e.g. lettuce, Lactuca sativa, and to detect
XX  PT resistance genes, e.g. to fingerprint cultivars
XX  PS Claim 37; Page 110; 183pp; English.
XX  CC The present sequence represents lettuce RG2A resistance protein,
XX  CC as deduced from the RG2A gene (see AAV44231) isolated from lettuce
XX  CC cultivar Diana genomic DNA. RG (or RUG) families RG1-RG5 and RG7
XX  CC have been identified, each RG family being defined as a group of
XX  CC polypeptide sequences that have at least 60% amino acid sequence
XX  CC identity. Individual members of an RG family typically map to the
XX  CC same genomic locus. RG nucleic acid sequences (see AAV44188-257) can
XX  CC be used to confer resistance in plants to a variety of pests,
XX  CC including viruses, fungi, nematodes, insects and bacteria.
XX  CC Sequences from within the RG genes can be used to fingerprint
XX  CC cultivars or germplasm for the presence of desired resistance
XX  CC genes. Antibodies to RG proteins (see AAW64150-98) are useful for
XX  CC detecting the type and amount of RG protein expressed in a plant
XX  sample.
XX  SQ Sequence 1890 AA;
XX  QY 1 QYLRALGK 9
XX  DB 31 QYRREMGK 39
XX  Query Match 75.0%; Score 33; DB 19; Length 1890;
XX  Best Local Similarity 55.6%; Pred. No. 9e+02;
XX  Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
XX  RESULT 12
XX  AA095491
XX  ID AA095491 standard; Protein: 1890 AA.
XX  AC AA095491;
XX  DT 02-JUL-2002 (first entry)
XX  DE Lettuce pest resistance protein RG2A #3.
XX  KW Plant; lettuce; resistance gene; plant downy mildew; RG; transgenic.
XX  KW Lactuca sativa.
XX  OS US6350933-B1.
XX  PN US6350933-B1.
XX  PD 26-FEB-2002.
XX  PF 09-JAN-1998; 98US-0004838.
XX  PR 10-JAN-1997; 97US-0781734.
XX  PA (REGC ) UNIV CALIFORNIA.
XX  PI Michelmore RW, Shen KA, Meyers BC;
XX  DR WPI: 2002-314681/35.

```



DR N-PSDB; ABK67843.  
XX Isolated nucleic acid molecule, RG2B (resistance gene 2B)  
PT polynucleotide, encodes polypeptide useful for conferring resistance to  
PT plant downy mildew  
XX  
PS Disclosure; Column 209-220; 209pp; English.  
CC The invention relates to an isolated nucleic acid molecule comprises an  
CC RG2B (resistance gene 2B) polynucleotide encoding a polypeptide  
CC conferring resistance to plant downy mildew, hybridises to a fully  
CC defined sequence of 15062 base pairs as given in the specification, under  
CC a wash with 0.2 x SSC (standard sodium citrate) at 65plusOC. Also  
CC included is a transgenic plant comprising a heterologous expression  
CC cassette containing a promoter operably linked to an RG2B polynucleotide  
CC encoding a polypeptide conferring resistance to plant downy mildew and  
CC hybridises to a fully defined sequence of 1323 amino acids as  
CC given in the specification, under a wash with 0.2 x SSC at 65plusOC.  
CC Also disclosed are other pest resistance genes and their protein  
CC products. The isolated nucleic acid molecule is used to confer resistance  
CC to downy mildew in a plant. This comprises introducing into the plant  
CC an expression cassette comprising a promoter operably linked to an RG2B  
CC polynucleotide. The present sequence is a lettuce resistance  
CC protein sequence (or protein consensus sequence).  
SQ Sequence 1890 AA:  
Query Match 75.0%; Score 33; DB 23; Length 1890;  
Best Local Similarity 55.6%; Pred. No. 9e+02; 1; Indels 0; Gaps 0;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
OY 1 QYLRALGLK 9  
||:|:|:|  
Db 31 QYRREKGIK 39  
RESULT 13  
ABK66978  
ID ABB66978 standard; Protein: 194 AA.  
XX  
AC ABB66978;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 27726.  
XX  
KW Drosophila: developmental biology; cell signalling; insecticide;  
XX pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PE 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
XX  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, LI PWD, Myers EW;  
XX  
DR WI: 2001-656860/75.  
XX  
DR N-PSDB; ABL11081.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Disclosure; SEQ ID NO 27726; 21pp + Sequence Listing; English.  
XX

CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABK01840-ABK16175) and the encoded proteins  
CC (ABK57737-ABK72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
SQ Sequence 194 AA:  
Query Match 72.7%; Score 32; DB 22; Length 194;  
Best Local Similarity 66.7%; Pred. No. 1.4e+02; 1; Indels 0; Gaps 0;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
OY 1 QYLRALGLK 9  
||:|:|:|  
Db 50 QYCRSLGLQ 58  
RESULT 14  
AAG70876  
ID AAG70876 standard; Protein: 281 AA.  
XX  
AC AAG70876;  
XX  
DT 27-JUL-2001 (first entry)  
XX  
DE C albicans apoptosis associated protein #56.  
XX  
KW Yeast; fungus; apoptosis; infection; proliferative disease;  
XX vaccine; autoimmune disease; ischaemia; neurodegeneration.  
XX  
OS Candida albicans.  
XX  
PN WO200102550-A2.  
XX  
PD 11-JAN-2001.  
XX  
PE 03-JUL-2000; 2000WO-BE00077.  
XX  
PR 01-JUL-1999; 99EP-0870141.  
XX  
PA (JANC ) JANSSEN PHARM NV.  
XX  
PI Contreras RH, De Backer MD, Luyten WHML, Malcorps IKL;  
XX Nelissen BJM, Reekmans RJ;  
XX  
DR WI: 2001-367042/38.  
XX  
DR N-PSDB; AAH29912.  
XX  
PT Yeast and fungal nucleic acids encoding proteins involved in a pathway  
PT leading to programmed cell death, useful for treating proliferative  
PT disorders; yeast and fungal infections, or for preventing apoptosis in  
PT certain diseases -  
XX  
PS Claim 24; Fig 2; 218pp; English.  
XX  
XX The present invention provides the protein and coding sequences of a  
CC number of apoptosis associated proteins from the yeast Saccharomyces  
CC cerevisiae and the fungus Candida albicans. These can be used to identify  
CC treatments for fungal and yeast infections, for proliferative diseases  
CC and for apoptosis related diseases such as autoimmune diseases, ischaemia  
CC and neurodegeneration. The present sequence is one of the C. albicans  
CC proteins of the invention.  
SQ Sequence 281 AA:  
Query Match 72.7%; Score 32; DB 22; Length 281;  
Best Local Similarity 75.0%; Pred. No. 2e+02;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 YLRALGK 9  
 ||| |||  
 Db 95 YLRALGK 102

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 QYRALGK 9  
 ||| |||  
 Db 313 EYRALGK 321

Search completed: July 11, 2003, 11:54:21  
 Job time : 6.16424 secs

RESULT 15  
 AAU35870  
 ID AAU35870 standard; Protein; 327 AA.

XX AAU35870;

AC 14-FEB-2002 (first entry)

DE Helicobacter pylori cellular proliferation protein #183.

KM Antisense; prokaryotic cellular proliferation protein;

XX antibiotic; antibacterial; drug design.

OS Helicobacter pylori.

PN WO200170955-A2.

PD 27-SEP-2001.

PF 21-MAR-2001; 2001WO-US09180.

PR 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207727P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

PI Yamamoto RT, Xu HH;

DR N-PSDB; AAS53729.

DR MPI: 2001-611495/70.

XX New polynucleotides for the identification and development of

PT antibiotics, comprise sequences of antisense nucleic acids -

PS Example 3; Seq ID No 11463; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to

CC prokaryotic cellular proliferation, their use in identifying the

CC genes, their use in the discovery of novel antibiotics, the essential

CC genes themselves and the encoded proteins. The prokaryotes used are

CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella

CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The

CC invention is also useful for the identification of potential new targets

CC for antibiotic development. The antisense nucleic acids can also be used

CC to identify proteins used in proliferation, to express these proteins,

CC and to obtain antibodies capable of binding to the expressed proteins.

CC The proteins can be used to screen compounds in rational drug discovery

CC programmes. The antisense nucleic acid sequence is also useful to screen

CC for homologous nucleic acids which are required for cell proliferation in

CC a wide variety of organisms. The present sequence represents an

CC essential prokaryotic cellular proliferation protein.

CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic

CC format directly from WIPO at

CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 327 AA:

Query Match 72.7%; Score 32; DB 22; Length 327;

Best Local Similarity 66.7%; Pred. No. 2.4e+02;

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:49:36 ; Search time 1.60915 Seconds  
(without alignments)  
164.563 Million cell updates/sec

Title: US-10-091-442-5  
Perfect score: 44  
Sequence: 1 OYLRAIGLK 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Issued\_Patents\_AA:\*

1: /cgn2\_6/p10data/1/1aa/5A.COMB.pep:.\*  
2: /cgn2\_6/p10data/1/1aa/5B.COMB.pep:.\*  
3: /cgn2\_6/p10data/1/1aa/6A.COMB.pep:.\*  
4: /cgn2\_6/p10data/1/1aa/6B.COMB.pep:.\*  
5: /cgn2\_6/p10data/1/1aa/PCTUS.COMB.pep:.\*  
6: /cgn2\_6/p10data/1/1aa/backfile1.pep:.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	100.0	9	2	US-08-611-977-5
2	44	100.0	380	2	US-08-472-659-34
3	44	100.0	380	2	US-08-474-661-34
4	44	100.0	380	2	US-08-611-977-34
5	39	88.6	9	2	US-08-472-659-5
6	39	88.6	9	2	US-08-474-661-5
7	33	75.0	250	4	US-09-167-717-2
8	33	75.0	1366	4	US-09-004-838-22
9	33	75.0	1890	4	US-09-004-838-88
10	31	70.5	387	4	US-09-314-847A-4
11	31	70.5	387	4	US-09-570-778A-2
12	31	70.5	387	4	US-09-570-778A-7
13	31	70.5	387	4	US-09-570-778A-8
14	31	70.5	387	4	US-09-570-778A-9
15	31	70.5	623	4	US-09-347-801-4
16	31	70.5	623	4	US-09-347-801-4
17	31	70.5	803	1	US-08-062-368-2
18	31	70.5	803	1	US-08-062-368-4
19	31	70.5	2548	4	US-09-172-422-1
20	30	68.2	223	4	US-08-896-410-4
21	30	68.2	310	4	US-09-071-035-412
22	30	68.2	327	4	US-08-896-410-38
23	30	68.2	347	4	US-09-071-035-410
24	30	68.2	387	3	US-08-968-563-37
25	30	68.2	387	4	US-08-969-583A-37
26	30	68.2	387	4	US-09-570-778A-10
27	30	68.2	387	4	US-09-570-778A-11

28	30	68.2	393	4	US-09-134-001C-5594	Sequence 5594, App
29	30	68.2	512	4	US-09-134-078-26	Sequence 26, Appl
30	30	68.2	533	4	US-09-360-197-10	Sequence 10, Appl
31	30	68.2	1327	4	US-09-196-387-2	Sequence 2, Appl1
32	29	65.9	10	1	US-08-300-386A-60	Sequence 60, Appl
33	29	65.9	10	3	US-08-931-645-60	Sequence 60, Appl
34	29	65.9	10	5	PCT-US94-01258-60	Sequence 60, Appl
35	29	65.9	10	5	PCT-US95-11235-60	Sequence 60, Appl
36	29	65.9	214	4	US-09-228-986-129	Sequence 129, App
37	29	65.9	223	4	US-08-928-941D-22	Sequence 22, Appl
38	29	65.9	223	4	US-09-280-590A-22	Sequence 22, Appl
39	29	65.9	315	4	US-09-439-261-15	Sequence 15, Appl
40	29	65.9	315	4	US-09-227-613-16	Sequence 16, Appl
41	29	65.9	349	2	US-07-857-224B-95	Sequence 95, Appl
42	29	65.9	372	4	US-08-928-941D-16	Sequence 16, Appl
43	29	65.9	372	4	US-08-280-590A-16	Sequence 16, Appl
44	29	65.9	441	4	US-09-287-937C-11	Sequence 11, Appl
45	29	65.9	446	2	US-08-833-610-2	Sequence 2, Appl1

#### ALIGNMENTS

RESULT 1  
US-08-611-977-5  
Sequence 5, Application US/08611977  
Patent No. 5972886  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
INVENTOR: IWASA, Fuyuki  
APPLICANT: TSURUOKA, No. 5972886uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju 5972886uhiro  
APPLICANT: ISHIDA, No. 5972886uhiro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAGUCHI, Kozo  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/611,977  
FILING DATE: 06-MAR-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-204  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:

LENGTH: 9 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-611-977-5

Query Match 100.0%; Score 44; DB 2; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.9e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYLRALGK 9  
DB 1 QYLRALGK 9

RESULT 2  
US-08-472-659-34  
Sequence 34, Application US/08472659  
Patent No. 5831030

GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSURUOKA, No. 5831030uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. 5831030uhiro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, No. 5831030omi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P. O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/472,659  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-248  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-472-659-34

Query Match 100.0%; Score 44; DB 2; Length 380;

Best Local Similarity 100.0%; Pred. No. 0.079;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYLRALGK 9  
DB 289 QYLRALGK 297

RESULT 3  
US-08-474-661-34  
Sequence 34, Application US/08474661  
Patent No. 5874253

GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSURUOKA, No. 5874253uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. 5874253uhiro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, No. 5874253omi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/474,661  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: REA, TERESA STANER  
REGISTRATION NUMBER: 30,427  
REFERENCE/DOCKET NUMBER: 001560-204  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-6620  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-474-661-34

Query Match 100.0%; Score 44; DB 2; Length 380;  
Best Local Similarity 100.0%; Pred. No. 0.079;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYLRALGK 9  
DB 289 QYLRALGK 297

RESULT 4

US-08-611-977-34  
: Sequence 34, Application US/08611977  
: Patent No. 5972886  
: GENERAL INFORMATION:  
: APPLICANT: TSUJIMOTO, Masafumi  
: APPLICANT: IWASA, Fuyuki  
: APPLICANT: TSUROUOKA, No. 5972886uo  
: APPLICANT: NAKAZATO, Hiroshi  
: APPLICANT: MIURA, Kenju  
: APPLICANT: ISHIDA, No. 5972886unhiro  
: APPLICANT: KURIHARA, Tatsuya  
: APPLICANT: YAMAGUCHI, Kozo  
: TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
: NUMBER OF SEQUENCES: 34  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Burns, Doane, Swecker & Mathis  
: STREET: P.O. Box 1404  
: CITY: Alexandria  
: STATE: Virginia  
: COUNTRY: United States  
: ZIP: 22313-1404  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patentin Release #1.0, Version #1.30  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/611.977  
: FILING DATE: 06-MAR-1996  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 08/091.028  
: FILING DATE: 14-JUL-1993  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: JP 4-212305  
: FILING DATE: 17-JUL-1992  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: JP 6-067339  
: FILING DATE: 04-MAR-1993  
: ATTORNEY/AGENT INFORMATION:  
: NAME: McGowan, Malcolm K.  
: REGISTRATION NUMBER: 39,300  
: REFERENCE/DOCKET NUMBER: 001560-204  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (703) 836-6620  
: TELEFAX: (703) 836-2021  
: INFORMATION FOR SEQ ID NO: 34:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 380 amino acids  
: TYPE: amino acid  
: TOPOLOGY: linear  
: MOLECULE TYPE: protein  
: US-08-611-977-34

Query Match 100.0%; Score 44; DB 2; Length 380;  
Best Local Similarity 100.0%; Pred. No. 0.079;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYLRALGLK 9  
DB 289 QYLRALGLK 297

RESULT 5  
US-08-472-659-5  
: Sequence 5, Application US/08472659  
: Patent No. 5831030  
: GENERAL INFORMATION:  
: APPLICANT: TSUJIMOTO, Masafumi  
: APPLICANT: IWASA, Fuyuki  
: APPLICANT: TSUROUOKA, No. 5831030uo  
: APPLICANT: NAKAZATO, Hiroshi

APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. 5831030unhiro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAGUCHI, Kozo  
APPLICANT: YAMAGUCHI, No. 5831030om1  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/472.659  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091.028  
FILING DATE: 14-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-248  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-472-659-5

Query Match 88.6%; Score 39; DB 2; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.9e+05;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 YLRALGLK 9  
DB 2 YLRALGLK 9

RESULT 6  
US-08-474-661-5  
: Sequence 5, Application US/08474661  
: Patent No. 5874253  
: GENERAL INFORMATION:  
: APPLICANT: TSUJIMOTO, Masafumi  
: APPLICANT: IWASA, Fuyuki  
: APPLICANT: TSUROUOKA, No. 5874253uo  
: APPLICANT: NAKAZATO, Hiroshi  
: APPLICANT: MIURA, Kenju  
: APPLICANT: ISHIDA, No. 5874253unhiro  
: APPLICANT: KURIHARA, Tatsuya  
: APPLICANT: YAMAGUCHI, Kozo  
: APPLICANT: YAMAGUCHI, No. 5874253om1  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34

;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Burns, Doane, Swecker & Mathis  
;; STREET: George Mason Bldg., Washington & Prince Sts.  
;; CITY: Alexandria  
;; STATE: Virginia  
;; COUNTRY: United States  
;; ZIP: 22313-1404  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patent Release #1.0, Version #1.25  
;;  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/474,661  
;; FILING DATE: 07-JUN-1995  
;; CLASSIFICATION: 435  
;; PRIORITY APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/091,028  
;; FILING DATE: 14-JUL-1993  
;; APPLICATION NUMBER: JP 4-212305  
;; FILING DATE: 17-JUL-1992  
;; PRIORITY APPLICATION DATA:  
;; APPLICATION NUMBER: JP 5-067339  
;; FILING DATE: 04-MAR-1993  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: REA, TERESA STANER  
;; REGISTRATION NUMBER: 30,427  
;; REFERENCE/DOCKET NUMBER: 001560-204  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (703) 836-6620  
;; TELEFAX: (703) 836-6620  
;; INFORMATION FOR SEQ ID NO: 5:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 9 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;;  
US-08-474-661-5

Query Match 88.6%; Score 39; DB 2; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.9e+05;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 YLRALGLK 9  
Db 2 YLRALGLK 9

RESULT 7  
US-09-167-717-2  
; Sequence 2, Application US/09167717  
; Patent No. 6133014  
; GENERAL INFORMATION:  
; APPLICANT: MUKOYAMA, MASAHARU  
; APPLICANT: YASUDA, SHINZO  
; APPLICANT: KOMATSUZAKI, SATOMI  
; TITLE OF INVENTION: MALEATE ISOMERASE GENE  
; FILE REFERENCE: PH-555US  
; CURRENT APPLICATION NUMBER: US/09/167,717  
; CURRENT FILING DATE: 1998-10-07  
; EARLIER APPLICATION NUMBER: JP 276261/1997  
; EARLIER FILING DATE: 1997-10-08  
; NUMBER OF SEQ ID NOS: 14  
; SEQ ID NO 2  
; LENGTH: 250  
; TYPE: PRT  
; ORGANISM: Citrobacter amalonaticus  
US-09-167-717-2

Query Match 75.0%; Score 33; DB 4; Length 250;  
Best Local Similarity 77.8%; Pred. No. 13;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 QYLRALGLK 9  
Db 228 QMKALGLK 236

RESULT 8  
US-09-004-838-22  
; Sequence 22, Application US/09004838  
; Patent No. 6350933  
; GENERAL INFORMATION:  
; APPLICANT: Michelmore, Richard W.  
; APPLICANT: Shen, Kathy  
; TITLE OF INVENTION: Procedures and Materials for  
; TITLE OF INVENTION: Confering Pest Resistance in Plants  
; NUMBER OF SEQUENCES: 140  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/004,838  
; FILING DATE: 09-JAN-1998  
; CLASSIFICATION: 800  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US 08/761,734  
; FILING DATE: 10-JAN-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Einhorn, Gregory P.  
; REGISTRATION NUMBER: 38,440  
; REFERENCE/DOCKET NUMBER: 023070-078810US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0300  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1366 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FEATURE:  
; NAME/KEY: -  
; LOCATION: 1..1366  
; OTHER INFORMATION: /note="RLG2a amino acids".  
US-09-004-838-22

Query Match 75.0%; Score 33; DB 4; Length 1366;  
Best Local Similarity 55.6%; Pred. No. 86;  
Matches 5; Conservative .3; Mismatches 1; Indels 0; Gaps 0;

QY 1 QYLRALGLK 9  
Db 31 QYRMWGIK 39

RESULT 9  
US-09-004-838-88  
; Sequence 88, Application US/09004838  
; Patent No. 6350933  
; GENERAL INFORMATION:  
; APPLICANT: Michelmore, Richard W.  
; APPLICANT: Shen, Kathy  
; APPLICANT: Meyers, Blake

TITLE OF INVENTION: Procedures and Materials for  
TITLE OF INVENTION: Confering Pest Resistance in Plants  
NUMBER OF SEQUENCES: 140  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/004,838  
FILING DATE: 09-JAN-1998  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/781,734  
FILING DATE: 10-JAN-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmhorn, Gregory P.  
REGISTRATION NUMBER: 38,440  
REFERENCE/DOCKET NUMBER: 023070-078810US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 88:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1890 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FEATURE:  
NAME/KEY:  
LOCATION: 1..1890  
OTHER INFORMATION: /note= "RC2A deduced sequence"  
US-09-004-838-88  
Query Match 75.0%; Score 33; DB 4; Length 1890;  
Best Local Similarity 55.6%; Pred. No. 1.2e+02;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
QY 1 QYLRALGLK 9  
DB 31 QYRREMGIK 39  
RESULT 10  
US-09-314-847A-4  
Sequence 4, Application US/09314847A  
Patent No. 6365410  
GENERAL INFORMATION:  
APPLICANT: Schellenberger, Volker  
APPLICANT: Liu, Amy D.  
APPLICANT: Selionova, Olga V.  
TITLE OF INVENTION: Directed Evolution of Microorganisms  
FILE REFERENCE: GC560  
CURRENT APPLICATION NUMBER: US/09/314,847A  
CURRENT FILING DATE: 2000-05-19  
NUMBER OF SEQ ID NOS: 15  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 4  
LENGTH: 387  
TYPE: PRT  
ORGANISM: E. blatte  
US-09-314-847A-4  
Query Match 70.5%; Score 31; DB 4; Length 387;  
Best Local Similarity 66.7%; Pred. No. 57;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 QYLRALGLK 9  
DB 342 QHLRELGVK 350  
RESULT 11  
US-09-570-778A-2  
Sequence 2, Application US/09570778A  
Patent No. 6468773  
GENERAL INFORMATION:  
APPLICANT: Trimbur, Donald E.  
APPLICANT: Whited, Gregory M.  
APPLICANT: Selionova, Olga V.  
TITLE OF INVENTION: Mutant 1,3-Propanediol Dehydrogenase  
FILE REFERENCE: GC580-2  
CURRENT APPLICATION NUMBER: US/09/570,778A  
CURRENT FILING DATE: 2000-05-14  
PRIOR APPLICATION NUMBER: US 60/134,868  
PRIOR FILING DATE: 1999-05-19  
NUMBER OF SEQ ID NOS: 12  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 387  
TYPE: PRT  
ORGANISM: Escherichia blataae  
US-09-570-778A-2  
Query Match 70.5%; Score 31; DB 4; Length 387;  
Best Local Similarity 66.7%; Pred. No. 57;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 QYLRALGLK 9  
DB 342 QHLRELGVK 350  
RESULT 12  
US-09-570-778A-7  
Sequence 7, Application US/09570778A  
Patent No. 6468773  
GENERAL INFORMATION:  
APPLICANT: Trimbur, Donald E.  
APPLICANT: Whited, Gregory M.  
APPLICANT: Selionova, Olga V.  
TITLE OF INVENTION: Mutant 1,3-Propanediol Dehydrogenase  
FILE REFERENCE: GC580-2  
CURRENT APPLICATION NUMBER: US/09/570,778A  
CURRENT FILING DATE: 2000-05-14  
PRIOR APPLICATION NUMBER: US 60/134,868  
PRIOR FILING DATE: 1999-05-19  
NUMBER OF SEQ ID NOS: 12  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 7  
LENGTH: 387  
TYPE: PRT  
ORGANISM: Escherichia blataae  
US-09-570-778A-7  
Query Match 70.5%; Score 31; DB 4; Length 387;  
Best Local Similarity 66.7%; Pred. No. 57;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 QYLRALGLK 9  
DB 342 QHLRELGVK 350  
RESULT 13  
US-09-570-778A-8  
Sequence 8, Application US/09570778A  
Patent No. 6468773

```

; GENERAL INFORMATION:
; APPLICANT: Trimbur, Donald E.
; APPLICANT: Whited, Gregory M.
; APPLICANT: Selfonova, Olga V.
; TITLE OF INVENTION: Mutant 1,3-Propanediol Dehydrogenase
; FILE REFERENCE: GC580-2
; CURRENT APPLICATION NUMBER: US/09/570,778A
; CURRENT FILING DATE: 2000-05-14
; PRIOR APPLICATION NUMBER: US 60/134,868
; PRIOR FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 387
; TYPE: PRF
; ORGANISM: Escherichia blattae
US-09-570-778A-8

```

```

Query Match          70.5%; Score 31; DB 4; Length 387;
Best Local Similarity 66.7%; Pred. No. 57;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 QYLRALGK 9
        1:11111:1
DB      342 QHLRELGVK 350

```

```

RESULT 14
US-09-570-778A-9
; Sequence 9, Application US/09570778A
; Patent No. 6468773
; GENERAL INFORMATION:
; APPLICANT: Trimbur, Donald E.
; APPLICANT: Whited, Gregory M.
; APPLICANT: Selfonova, Olga V.
; TITLE OF INVENTION: Mutant 1,3-Propanediol Dehydrogenase
; FILE REFERENCE: GC580-2
; CURRENT APPLICATION NUMBER: US/09/570,778A
; CURRENT FILING DATE: 2000-05-14
; PRIOR APPLICATION NUMBER: US 60/134,868
; PRIOR FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 387
; TYPE: PRF
; ORGANISM: Escherichia blattae
US-09-570-778A-9

```

```

Query Match          70.5%; Score 31; DB 4; Length 387;
Best Local Similarity 66.7%; Pred. No. 57;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 QYLRALGK 9
        1:11111:1
DB      342 QHLRELGVK 350

```

```

RESULT 15
US-09-347-801-4
; Sequence 4, Application US/09347801
; Patent No. 6262345
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Lee, Jian Ming
; TITLE OF INVENTION: Plant Protein Kinases
; FILE REFERENCE: BB-1171
; CURRENT APPLICATION NUMBER: US/09/347,801
; CURRENT FILING DATE: 1999-07-02
; EARLIER APPLICATION NUMBER: 60/092,438
; EARLIER FILING DATE: JULY 10, 1998
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: Microsoft Office 97

```

```

; SEQ ID NO 4
; LENGTH: 623
; TYPE: PRF
; ORGANISM: Oryza sativa
US-09-347-801-4

```

```

Query Match          70.5%; Score 31; DB 4; Length 623;
Best Local Similarity 77.8%; Pred. No. 97;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      1 QYLRALGK 9
        11111111
DB      453 QYLRATPLK 461

```

```

Search completed: July 11, 2003, 12:02:08
Job time : 2.60915 secs

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: July 11, 2003, 11:59:17 ; Search time 2.63825 Seconds

(without alignments)  
397.148 Million cell updates/sec

Title: US-10-091-442-5

Perfect score: 44

Sequence: 1 QYLRALGLK 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 445758 seqs, 116419773 residues

Total number of hits satisfying chosen parameters: 445758

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

## Database :

Published\_Applications\_AA:\*

- 1: /cgn2\_6/ptodata/1/pubppaa/US07\_NEW\_PUB.pep:\*
- 2: /cgn2\_6/ptodata/1/pubppaa/PCR\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/1/pubppaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/1/pubppaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/1/pubppaa/PCTUS\_PUBCOMB.pep:\*
- 6: /cgn2\_6/ptodata/1/pubppaa/US07\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/1/pubppaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/1/pubppaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/1/pubppaa/US08\_NEW\_PUB.pep:\*
- 10: /cgn2\_6/ptodata/1/pubppaa/US09\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/1/pubppaa/US10\_NEW\_PUB.pep:\*
- 12: /cgn2\_6/ptodata/1/pubppaa/US10\_PUBCOMB.pep:\*
- 13: /cgn2\_6/ptodata/1/pubppaa/US60\_NEW\_PUB.pep:\*
- 14: /cgn2\_6/ptodata/1/pubppaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	100.0	9	US-10-091-442-5	Sequence 5, Appli
2	44	100.0	9	US-09-140-719-5	Sequence 5, Appli
3	44	100.0	380	US-10-091-442-34	Sequence 34, Appli
4	44	100.0	380	US-09-140-719-34	Sequence 34, Appli
5	34	77.3	350	US-09-847-208-53	Sequence 53, Appli
6	33	75.0	1465	US-10-083-357-1310	Sequence 1310, Ap
7	32	72.7	327	US-09-815-242-1163	Sequence 1163, A
8	32	72.7	327	US-09-815-242-11625	Sequence 11625, A
9	32	72.7	342	US-09-815-242-12104	Sequence 12104, A
10	31	70.5	292	US-10-023-282-1116	Sequence 1116, Ap
11	31	70.5	387	US-10-037-677-4	Sequence 4, Appli
12	31	70.5	387	US-09-991-138-2	Sequence 2, Appli
13	31	70.5	387	US-09-991-138-7	Sequence 7, Appli
14	31	70.5	387	US-09-991-138-8	Sequence 8, Appli
15	31	70.5	387	US-09-991-138-9	Sequence 9, Appli
16	31	70.5	424	US-10-179-784-3	Sequence 3, Appli
17	31	70.5	444	US-10-156-761-10914	Sequence 10914, A
18	31	70.5	478	US-10-128-714-3188	Sequence 3188, Ap
19	31	70.5	478	US-10-128-714-8188	Sequence 8188, Ap

20	31	70.5	623	US-09-854-731-4	Sequence 4, Appli
21	31	70.5	625	US-09-854-731-18	Sequence 18, Appli
22	31	70.5	905	US-10-156-761-14694	Sequence 14694, A
23	31	70.5	2548	US-09-851-682A-1	Sequence 1, Appli
24	30	68.2	328	US-09-895-913A-134	Sequence 134, App
25	30	68.2	331	US-10-147-874-2	Sequence 2, Appli
26	30	68.2	387	US-09-308-207-37	Sequence 37, Appli
27	30	68.2	387	US-09-991-138-10	Sequence 10, Appli
28	30	68.2	387	US-09-991-138-11	Sequence 11, Appli
29	30	68.2	449	US-09-815-242-11349	Sequence 11349, A
30	30	68.2	512	US-10-121-032-26	Sequence 26, Appli
31	30	68.2	512	US-10-093-037-26	Sequence 26, Appli
32	30	68.2	752	US-09-764-853-641	Sequence 641, App
33	30	68.2	1094	US-10-156-761-11435	Sequence 11435, A
34	30	68.2	1291	US-10-136-761-14161	Sequence 14161, A
35	30	68.2	1327	US-09-972-115A-8	Sequence 8, Appli
36	30	68.2	1327	US-09-841-835-2	Sequence 2, Appli
37	29	65.9	60	US-10-156-761-12478	Sequence 12478, A
38	29	65.9	61	US-09-738-626-4102	Sequence 4102, Ap
39	29	65.9	148	US-10-101-464A-553	Sequence 553, App
40	29	65.9	149	US-09-764-864-1171	Sequence 1171, Ap
41	29	65.9	214	US-10-101-464A-129	Sequence 129, App
42	29	65.9	214	US-09-864-636A-164	Sequence 164, App
43	29	65.9	223	US-09-892-398-22	Sequence 22, Appli
44	29	65.9	315	US-10-191-513A-16	Sequence 16, Appli
45	29	65.9	339	US-09-815-242-5446	Sequence 5446, Ap

## ALIGNMENTS

RESULT 1  
US-10-091-442-5  
Sequence 5, Application US/10091442  
Patent No. US20020164711A1  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
INVENTOR: IWASA, Fuyuki  
TSURUDOKA, No. US20020164711A1  
NAKAZATO, Hiroshi  
MIURA, Kenji  
ISHIDA, No. US20020164711A1  
KURIHARA, Tatsuya  
YAMACHI, KOZO  
YAMAGUCHI, No. US20020164711A1  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/091,442  
FILING DATE: 07-Mar-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
APPLICATION NUMBER: JP 6-067339

FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
US-10-091-442-5

Query Match 100.0%; Score 44; DB 9; Length 9;  
Best Local Similarity 100.0%; Pred. No. 4e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYLRALGLK 9  
Db 1 QYLRALGLK 9

RESULT 2  
US-09-140-719-5  
Sequence 5, Application US/09140719  
Patent No. US20010026931A1  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSUROJOKA, No. US20010026931A1uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. US20010026931A1uhiro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAGUCHI, Kozo  
APPLICANT: YAMAGUCHI, No. US20010026931A1omi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.

REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-140-719-5

Query Match 100.0%; Score 44; DB 10; Length 9;  
Best Local Similarity 100.0%; Pred. No. 4e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYLRALGLK 9  
Db 1 QYLRALGLK 9

RESULT 3  
US-10-091-442-34  
Sequence 34, Application US/10091442  
Patent No. US20020164711A1  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSUROJOKA, No. US20020164711A1uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. US20020164711A1uhiro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAGUCHI, Kozo  
APPLICANT: YAMAGUCHI, No. US20020164711A1omi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/091,442  
FILING DATE: 07-MAR-2002  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021

INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 34:  
US-10-091-442-34

Query Match 100.0%; Score 44; DB 9; Length 380;  
Best Local Similarity 100.0%; Pred. No. 0.8;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYLRALGK 9  
DB 289 QYLRALGK 297

RESULT 4  
US-09-140-719-34  
Sequence 34, Application US/09140719  
Patent No. US20010026931A1  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IMASA, Fuyuki  
APPLICANT: TSUBOIJOA, No. US20010026931A1uo  
APPLICANT: MAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. US20010026931A1uhltro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAGUCHI, Kozo  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
City: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
APPLICATION DATA:  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid

TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-140-719-34

Query Match 100.0%; Score 44; DB 10; Length 380;  
Best Local Similarity 100.0%; Pred. No. 0.8;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYLRALGK 9  
DB 289 QYLRALGK 297

RESULT 5  
US-09-847-208-53  
Sequence 53, Application US/09847208  
Publication No. US20030082190A1  
GENERAL INFORMATION:  
APPLICANT: Saxon, Andrew  
APPLICANT: Zhang, Ke  
APPLICANT: Zhu, Daocheng  
TITLE OF INVENTION: FUSION MOLECULES AND TREATMENT OF  
TITLE OF INVENTION: IGE-MEDIATED ALLERGIC DISEASES  
FILE REFERENCE: UC67.002A  
CURRENT APPLICATION NUMBER: US/09/847,208  
CURRENT FILING DATE: 2001-05-01  
NUMBER OF SEQ ID NOS: 177  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 53  
LENGTH: 350  
TYPE: PRT  
ORGANISM: Candida albicans (Yeast)  
US-09-847-208-53

Query Match 77.3%; Score 34; DB 9; Length 350;  
Best Local Similarity 66.7%; Pred. No. 66;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 QYLRALGK 9  
DB 191 QYLRALGK 199

RESULT 6  
US-10-083-357-1310  
Sequence 1310, Application US/10083357  
Publication No. US20030054370A1  
GENERAL INFORMATION:  
APPLICANT: Qianlong Zeng et al.  
TITLE OF INVENTION: Systemic Discovery of New Genes  
FILE REFERENCE: 032796-090  
CURRENT APPLICATION NUMBER: US/10/083,357  
CURRENT FILING DATE: 2002-02-27  
NUMBER OF SEQ ID NOS: 1346  
SEQ ID NO 1310  
LENGTH: 1465  
TYPE: PRT  
ORGANISM: Saccharomyces cerevisiae  
US-10-083-357-1310

Query Match 75.0%; Score 33; DB 9; Length 1465;  
Best Local Similarity 66.7%; Pred. No. 4.6e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 QYLRALGK 9  
DB 1094 QYLRALGK 1102

RESULT 7  
US-09-815-242-11463  
Sequence 11463, Application US/09815242  
Patent No. US20020061569A1

```

; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 11463
; LENGTH: 327
; TYPE: PRT
; ORGANISM: Helicobacter pylori
;
US-09-815-242-11463

Query Match
Best Local Similarity 72.7%; Score 32; DB 10; Length 327;
Pred. No. 1.5e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 QYLRALGLK 9
DB 313 EYLRNALGLQ 321

RESULT 8
US-09-815-242-11625
; Sequence 11625, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 12104
; LENGTH: 342
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
;
US-09-815-242-12104

Query Match
Best Local Similarity 72.7%; Score 32; DB 10; Length 342;
Pred. No. 1.5e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 11625
; LENGTH: 327
; TYPE: PRT
; ORGANISM: Helicobacter pylori
;
US-09-815-242-11625

Query Match
Best Local Similarity 72.7%; Score 32; DB 10; Length 327;
Pred. No. 1.5e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 QYLRALGLK 9
DB 313 EYLRNALGLQ 321

RESULT 9
US-09-815-242-12104
; Sequence 12104, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 12104
; LENGTH: 342
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
;
US-09-815-242-12104

Query Match
Best Local Similarity 72.7%; Score 32; DB 10; Length 342;
Pred. No. 1.5e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 QYLRALGL 8
DB 188 QYARAMGL 195

RESULT 10
US-10-023-282-1116
; Sequence 1116, Application US/10023282

```

```
Publication No. US20030092893A1
GENERAL INFORMATION:
APPLICANT: Young et al.
TITLE OF INVENTION: 207 Human Secreted Proteins
FILE REFERENCE: p7007P1
CURRENT APPLICATION NUMBER: US/10/023,282
EARLIER FILING DATE: 2001-12-20
EARLIER APPLICATION NUMBER: 09/205,258
EARLIER FILING DATE: 1998-12-04
EARLIER APPLICATION NUMBER: PCT/US98/11422
EARLIER FILING DATE: 1998-06-04
EARLIER APPLICATION NUMBER: 60/048,885
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049,375
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,881
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,880
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,896
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049,020
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,876
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,895
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,884
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,894
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,971
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,964
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,882
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,899
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,893
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,900
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,901
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,892
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,915
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049,019
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,970
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,972
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,916
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049,374
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049,373
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,875
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049,374
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,917
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,949
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,974
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,883
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,897
EARLIER FILING DATE: 1997-06-06

EARLIER APPLICATION NUMBER: 60/048,898
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,962
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,963
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,877
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,878
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/070,923
EARLIER FILING DATE: 1997-12-18
EARLIER APPLICATION NUMBER: 60/092,921
EARLIER FILING DATE: 1998-07-15
EARLIER APPLICATION NUMBER: 60/094,657
EARLIER FILING DATE: 1998-07-30
NUMBER OF SEQ ID NOS: 1227
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1116
LENGTH: 292
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (11)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
NAME/KEY: SITE
LOCATION: (15)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
NAME/KEY: SITE
LOCATION: (35)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
NAME/KEY: SITE
LOCATION: (36)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
NAME/KEY: SITE
LOCATION: (39)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
NAME/KEY: SITE
LOCATION: (40)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
NAME/KEY: SITE
LOCATION: (45)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
NAME/KEY: SITE
LOCATION: (91)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
NAME/KEY: SITE
LOCATION: (255)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
NAME/KEY: SITE
LOCATION: (256)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
NAME/KEY: SITE
LOCATION: (257)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
NAME/KEY: SITE
LOCATION: (258)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-023-282-1116

Query Match          70.5%; Score 31; DB 9; Length 292;
```

Best Local Similarity 85.7%; Pred. No. 2.1e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 OYLRALGK 7  
1:11111  
DB 280 QHLRELGVK 286

## RESULT 11

US-10-037-677-4  
; Sequence 4, Application US/10037677  
; Patent No. US20020173003A1  
; GENERAL INFORMATION:  
; APPLICANT: Schellenberger, Volker  
; APPLICANT: Liu, Amy D.  
; APPLICANT: Selifonova, Olga V.  
; TITLE OF INVENTION: Directed Evolution of Microorganisms  
; FILE REFERENCE: GC560  
; CURRENT APPLICATION NUMBER: US/10/037,677  
; CURRENT FILING DATE: 2001-10-23  
; PRIOR APPLICATION NUMBER: 09/314,847  
; PRIOR FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4  
; LENGTH: 387  
; TYPE: PRT  
; ORGANISM: E. blatte  
US-10-037-677-4

Query Match 70.5%; Score 31; DB 9; Length 387;  
Best Local Similarity 66.7%; Pred. No. 2.8e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 OYLRALGK 9  
1:11111  
DB 342 QHLRELGVK 350

## RESULT 12

US-09-991-138-2  
; Sequence 2, Application US/09991138  
; Publication No. US20030040091A1  
; GENERAL INFORMATION:  
; APPLICANT: Trimbur, Donald E.  
; APPLICANT: Whited, Gregory M.  
; APPLICANT: Selifonova, Olga V.  
; TITLE OF INVENTION: Mutant 1,3-Propanediol Dehydrogenase  
; FILE REFERENCE: GC580-2D1  
; CURRENT APPLICATION NUMBER: US/09/991,138  
; CURRENT FILING DATE: 2001-11-16  
; PRIOR APPLICATION NUMBER: US 09/570,778  
; PRIOR FILING DATE: 2000-05-14  
; PRIOR APPLICATION NUMBER: US 60/134,868  
; PRIOR FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 387  
; TYPE: PRT  
; ORGANISM: Escherichia blattae  
US-09-991-138-2

Query Match 70.5%; Score 31; DB 9; Length 387;  
Best Local Similarity 66.7%; Pred. No. 2.8e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 OYLRALGK 9  
1:11111  
DB 342 QHLRELGVK 350

## RESULT 13

US-09-991-138-7  
; Sequence 7, Application US/09991138  
; Publication No. US20030040091A1  
; GENERAL INFORMATION:  
; APPLICANT: Trimbur, Donald E.  
; APPLICANT: Whited, Gregory M.  
; APPLICANT: Selifonova, Olga V.  
; TITLE OF INVENTION: Mutant 1,3-Propanediol Dehydrogenase  
; FILE REFERENCE: GC580-2D1  
; CURRENT APPLICATION NUMBER: US/09/991,138  
; CURRENT FILING DATE: 2001-11-16  
; PRIOR APPLICATION NUMBER: US 09/570,778  
; PRIOR FILING DATE: 2000-05-14  
; PRIOR APPLICATION NUMBER: US 60/134,868  
; PRIOR FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7  
; LENGTH: 387  
; TYPE: PRT  
; ORGANISM: Escherichia blattae  
US-09-991-138-7

Query Match 70.5%; Score 31; DB 9; Length 387;  
Best Local Similarity 66.7%; Pred. No. 2.8e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 OYLRALGK 9  
1:11111  
DB 342 QHLRELGVK 350

## RESULT 14

US-09-991-138-8  
; Sequence 8, Application US/09991138  
; Publication No. US20030040091A1  
; GENERAL INFORMATION:  
; APPLICANT: Trimbur, Donald E.  
; APPLICANT: Whited, Gregory M.  
; APPLICANT: Selifonova, Olga V.  
; TITLE OF INVENTION: Mutant 1,3-Propanediol Dehydrogenase  
; FILE REFERENCE: GC580-2D1  
; CURRENT APPLICATION NUMBER: US/09/991,138  
; CURRENT FILING DATE: 2001-11-16  
; PRIOR APPLICATION NUMBER: US 09/570,778  
; PRIOR FILING DATE: 2000-05-14  
; PRIOR APPLICATION NUMBER: US 60/134,868  
; PRIOR FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 387  
; TYPE: PRT  
; ORGANISM: Escherichia blattae  
US-09-991-138-8

Query Match 70.5%; Score 31; DB 9; Length 387;  
Best Local Similarity 66.7%; Pred. No. 2.8e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 OYLRALGK 9  
1:11111  
DB 342 QHLRELGVK 350

## RESULT 15

US-09-991-138-9  
; Sequence 9, Application US/09991138  
; Publication No. US20030040091A1  
; GENERAL INFORMATION:  
; APPLICANT: Trimbur, Donald E.  
; APPLICANT: Whited, Gregory M.  
; APPLICANT: Selifonova, Olga V.

; TITLE OF INVENTION: Mutant 1,3-Propanediol Dehydrogenase  
; FILE REFERENCE: GC580-2D1  
; CURRENT APPLICATION NUMBER: US/09/991,138  
; CURRENT FILING DATE: 2001-11-16  
; PRIOR APPLICATION NUMBER: US 09/570,778  
; PRIOR FILING DATE: 2000-05-14  
; PRIOR APPLICATION NUMBER: US 60/134,868  
; PRIOR FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9  
; LENGTH: 387  
; TYPE: PRT  
; ORGANISM: Escherichia blattae  
; US-09-991-138-9

Query Match 70.5%; Score 31; DB 9; Length 387;  
Best Local Similarly 66.7%; Pred. No. 2.8e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Caps 0;

QY 1 QYLRALGLK 9  
1:11111:1  
Db 342 QHLRELGVK 350

Search completed: July 11, 2003, 12:37:33  
Job time : 2.63825 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: July 11, 2003, 11:45:06 ; Search time 1.81497 Seconds  
(without alignments)  
476.708 Million cell updates/sec

Title: US-10-091-442-5

Perfect score: 44

Sequence: 1 OYRALGLK 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

1: PIR73:\*  
2: PIR1:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35	79.5	164	2 I40526	hypothetical prote
2	35	79.5	270	2 B69781	multidrug-efflux t
3	34	77.3	80	2 F72303	glutaredoxin - The
4	34	77.3	187	2 T06581	probable deoxyribo
5	34	77.3	350	1 S52153	alcohol dehydrogen
6	34	77.3	406	2 C75537	conserved hypotet
7	33	75.0	127	2 B64486	hypothetical prote
8	33	75.0	412	2 B81356	ankyrin repeat-con
9	33	75.0	413	2 E70661	probable PE protei
10	33	75.0	1139	2 T08421	resistance protein
11	33	75.0	331262	2 S31262	TYB protein - yeas
12	33	75.0	1467	2 PC1253	TYB protein - yeas
13	33	75.0	1802	2 S52611	TYB protein - yeas
14	33	75.0	1803	2 S56894	TYB protein - yeas
15	32	72.7	239	2 E84332	hypothetical prote
16	32	72.7	261	2 S20610	calpastatin - mous
17	32	72.7	274	2 S59378	hypothetical prote
18	32	72.7	309	2 C83886	hypothetical prote
19	32	72.7	337	2 H64716	ABC transporter, A
20	32	72.7	337	2 F71800	ABC transporter, A
21	32	72.7	342	2 A82968	alcohol dehydrogen
22	32	72.7	342	2 T34045	hypothetical prote
23	32	72.7	373	2 AD3470	alcohol dehydrogen
24	32	72.7	408	2 T33939	probable alcohol d
25	32	72.7	434	2 B64376	hypothetical prote
26	32	72.7	447	2 A72358	conserved hypotet
27	32	72.7	487	2 AD0818	probable exported
28	32	72.7	509	2 JC5651	N-acetylglucosamin
29	32	72.7	516	2 T24644	hypothetical prote

30	32	72.7	523	2 F82254	PTS system, N-acet
31	32	72.7	603	2 S15074	calpastatin - rat
32	32	72.7	625	2 G96976	probable periplasm
33	32	72.7	654	2 T10772	calpastatin - rat
34	32	72.7	1150	2 T40395	lim domain protein
35	32	72.7	1740	2 T43215	ribonucleotide red
36	31	70.5	107	2 F72059	hypothetical prote
37	31	70.5	107	2 E86565	hypothetical prote
38	31	70.5	127	2 T17431	hypothetical prote
39	31	70.5	240	2 T35972	hypothetical prote
40	31	70.5	342	2 A97201	hemin permease (lm
41	31	70.5	349	2 A97124	RECA recombinase,
42	31	70.5	352	2 A24648	alcohol dehydrogen
43	31	70.5	359	2 T36443	probable penicilli
44	31	70.5	375	2 AG2565	hypothetical prote
45	31	70.5	380	2 H82566	gmh protein Xp236

## ALIGNMENTS

RESULT 1  
I40526  
hypothetical protein ypoC - Bacillus subtilis  
N:Alternate names: hypothetical protein y (ponA operon); jooC protein  
C:Species: Bacillus subtilis  
C>Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text-change 20-Jun-2000  
C/Accession: I40526; I40530; E69939  
R:Brund, C.; Sorokin, A.; Serron, P.; Ehrlich, S.D.  
Microbiology 141, 321-322, 1995  
A>Title: Nucleotide sequence of the Bacillus subtilis dnaD gene.  
A:Reference number: I40523; MUID:95219085; PMID:7704260  
A:Accession: I40526  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-164 <RES>  
A/Cross-references: EMBL:U11289; NID:9533096; PIDN:AAA0006.1; PID:9533100  
R:Popham, D.L.; Setlow, P.  
J. Bacteriol. 177, 326-335, 1995  
A>Title: Cloning, nucleotide sequence, and mutagenesis of the Bacillus subtilis ponA  
A:Reference number: I40527; MUID:95113769; PMID:7814321  
A:Accession: I40530  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-164 <RES>  
A/Cross-references: EMBL:U11883; NID:9520536; PIDN:AAA64948.1; PID:9520542  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber  
C.: Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;  
A.: Ehrlich, S.D.; Emmerston, P.T.; Ertlan, K.D.; Erlington, J.; Fabret, C.; Ferraril,  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Filtz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galliz, A.; Gal  
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Huilo, M  
Koeter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau  
Y, M.; Ogawa, K.; Ogihara, A.; Oudeg, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete  
Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scari  
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Selguchi, J.; Sekowska, A.; Se  
akenchu, M.; Tamakoshi, A.; Tanaka, T.; Terstata, P.; Tognoni, A.; Toseto, V.; Uchiya  
T.; Winters, P.; Wipal, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida  
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A>Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis  
A:Reference number: A69580; MUID:98044033; PMID:9384377  
A:Accession: E69939  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-164 <KUN>  
A/Cross-references: GB:Z99115; GB:AL009126; NID:92634478; PIDN:CAB14149.1; PID:926346  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: ypoC; jooC  
C:Superfamily: Bacillus subtilis hypothetical protein ypoC  
Query Match 79.5%; Score 35; DB 2; Length 164;

Best Local Similarity 77.8%; Pred. No. 8.9;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 QYLRALGLK 9  
Db 13 QYLRSLGOK 21

## RESULT 2

B69781

multidrug-efflux transporter regulator homolog ydfl - Bacillus subtilis

C:Species: Bacillus subtilis

C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 01-Mar-2002

C:Accession: B69781

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter, C.; Bron, S.; Brunielle, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall, lech, J.; Harwood, C.R.; Hentaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.

Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A.; Authors: Lauder, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, Y., M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot, akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipit, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K

A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A:Reference number: A69580; MUID:98044033; PMID:9384377

A:Accession: B69781

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-270 <KUN>

A:Cross-references: GB:299106; GB:AL009126; NID:g2632653; PIDN:CAB12353.1; PID:e1182512;

A:Experimental source: strain 168

C:Genetics:

A:Gene: ydfl

C:Superfamily: transcription activator bmr

Query Match 79.5%; Score 35; DB 2; Length 270;  
Best Local Similarity 87.5%; Pred. No. 15;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYLRALGL 8  
Db 56 KYLRALGL 63

RESULT 3  
F72303

glutaredoxin - Thermotoga maritima (strain MS8)

C:Species: Thermotoga maritima

C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000

C:Accession: F72303

R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; C.M.

Nature 399, 323-329, 1999

A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq

A:Reference number: A72200; MUID:99287316; PMID:10360571

A:Accession: F72303

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-80 <ARN>

A:Cross-references: GB:AE001764; GB:AE000512; NID:g4981561; PIDN:AMD36108.1; PID:g498157

A:Experimental source: strain MS8

C:Genetics:

A:Gene: TM1031

C:Superfamily: Clostridium pasteurianum probable glutaredoxin

C:Keywords: electron transfer

Query Match 77.3%; Score 34; DB 2; Length 80;

Best Local Similarity 66.7%; Pred. No. 7;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 QYLRALGLK 9  
Db 22 EYFRSLGOK 30

## RESULT 4

T06581

Probable deoxyribodipyrimidine photo-lyase (EC 4.1.99.3) - tomato (fragment)

N:Alternate names: cryptochrome 2; DNA photolyase; photoreactivating enzyme

C:Species: Lycopersicon esculentum (tomato)

C>Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 20-Jun-2000

C:Accession: T06581

R:Kojimaoglu, U.; Schneider-Poetsch, H.A. submitted to the EMBL data library, August 1997

A:Description: The evolution of cryptochrome genes.

A:Reference number: Z15773

A:Accession: T06581

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-187 <KOL>

A:Cross-references: EMBL:AJ000695; PIDN:CAA04247.1

C:Genetics:

A:Gene: CYP2

A:Note: Intron positions not resolved (incomplete sequence)

C:Superfamily: deoxyribodipyrimidine photo-lyase

C:Keywords: carbon-carbon lyase

Query Match 77.3%; Score 34; DB 2; Length 187;  
Best Local Similarity 75.0%; Pred. No. 16;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 YLRALGLK 9  
Db 147 YLRALGLR 154

## RESULT 5

S52153

alcohol dehydrogenase (EC 1.1.1.1) - yeast (Candida albicans)

N:Alternate names: 40K allergen

C:Species: Candida albicans

C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 08-Dec-2000

C:Accession: S63781; A61504; S52153

R:Berram, G.; Swoboda, R.K.; Gooday, G.W.; Gow, N.A.R.; Brown, A.J.P.

Yeast 12, 115-127, 1996

A:Title: Structure and regulation of the Candida albicans ADH1 gene encoding an immun

A:Reference number: S63781; MUID:96287648; PMID:8686375

A:Accession: S63781

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-350 <BE2>

A:Cross-references: EMBL:X01694; NID:g608689; PIDN:CAA57342.1; PID:g608680

A:Note: only a part of the coding sequence is given

R:Shen, H.D.; Choo, K.B.; Lee, H.H.; Hsieh, J.C.; Lin, W.L.; Lee, W.R.; Han, S.H.

Clin. Exp. Allergy 21, 675-681, 1991

A:Title: The 40-kilodalton allergen of Candida albicans is an alcohol dehydrogenase:

A:Reference number: A61504; MUID:92136159; PMID:1777830

A:Accession: A61504

A:Molecule type: mRNA

A:Residues: 104-139, 'T', 'A', '228-312, 'S' <SHE>

A:Cross-references: GB:X81694; NID:g608689

A:Note: sequence extracted from NCBI backbone (NCBIN:80403, NCBIIP:80409)

C:Genetics:

A:Gene: ADH1

C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

C:Keywords: alcohol metabolism; NAD: oxidoreductase; zinc

F:31-338/Domain: long-chain alcohol dehydrogenase homology <LADH>

F:46,69,156/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted

Query Match 77.3%; Score 34; DB 1; Length 350;

Best Local Similarity 66.7%; Pred. No. 31;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 1 OYLRALGLK 9  
Db 191 OYLRALGLR 199

## RESULT 6

C75537

conserved hypothetical protein - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000

C:Accession: C75537

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1;

A:Reference number: A75250; MUID:20036896; PMID:10567266

A:Accession: C75537

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-406 &lt;WHI&gt;

A:Cross-references: GB:AE001890; GB:AE000513; NID:96457959; PIDN:AF09871.1; PID:9645796

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR0290

A:Map position: 1

C:Superfamily: tetracycline resistance protein

Query Match 77.3%; Score 34; DB 2; Length 406;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 YLRALGL 8  
Db 33 YLRALGL 39

## RESULT 7

B64486

hypothetical protein MJ1491 - Methanococcus jannaschii

C:Species: Methanococcus jannaschii

C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 21-Jul-2000

C:Accession: B64486

R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,

Reich, C.I.; Overbeek, R.; Kirsch, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;

rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.

Science 273, 1058-1073, 1996

A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C

A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii

A:Reference number: A64300; MUID:96337999; PMID:8688087

A:Accession: B64486

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-127 &lt;BUU&gt;

A:Cross-references: GB:U67590; GB:L77117; NID:91592126; PIDN:AAB99510.1; PID:91592128; T

C:Genetics:

A:Map position: FOR1465677-1466060

A:Start codon: GTC

C:Superfamily: Methanococcus jannaschii hypothetical protein MJ1491

Query Match 75.0%; Score 33; DB 2; Length 127;  
Best Local Similarity 87.5%; Pred. No. 18;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2 YLRALGLK 9  
Db 83 YLRALGLK 90

## RESULT 8

B81356  
ankyrin repeat-containing probable periplasmic protein Cj0834c [imported] - Campyloba

C:Species: Campylobacter jejuni

C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002

C:Accession: B81356

R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelsey, J.M.; Churcher, C.; Basham, D.; Chai

C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Bar

Nature 403, 665-668, 2000

A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals

A:Reference number: A81250; MUID:20150912; PMID:10688204

A:Accession: B81356

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-412 &lt;PAR&gt;

A:Cross-references: GB:A1139076; GB:A1111168; NID:96968128; PIDN:CAB73099.1; PID:9696

A:Experimental source: serotype O2, strain NCTC 11168

C:Genetics:

A:Gene: Cj0834c

Query Match 75.0%; Score 33; DB 2; Length 412;  
Best Local Similarity 75.0%; Pred. No. 58;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy 2 YLRALGLK 9  
Db 393 YLRALGLK 400

## RESULT 9

E70661

probable PE protein - Mycobacterium tuberculosis (strain H37Rv)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000

C:Accession: E70661

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon

Connor, R.; Davies, R.; Devlin, K.; Gentles, S.; Hamlin, N.; Holtroyd, Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno

A:Reference number: A70500; MUID:96295987; PMID:9634230

A:Accession: E70661

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-413 &lt;COL&gt;

A:Cross-references: GB:Z83860; GB:AL123456; NID:93261681; PIDN:CAB06157.1; PID:917812

A:Experimental source: strain H37Rv

C:Genetics:

A:Gene: PE

C:Superfamily: Phaeosolus glycine-rich cell wall protein 1.8

Query Match 75.0%; Score 33; DB 2; Length 413;  
Best Local Similarity 77.8%; Pred. No. 58;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 OYLRALGLK 9  
Db 400 OYLRALGLK 408

## RESULT 10

T08421

resistance protein homolog RGC2a - garden lettuce

N:Alternate names: resistance protein candidate

C:Species: Lactuca sativa (garden lettuce)

C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 11-Jun-1999

C:Accession: T08421

R:Shen, K.A.; Meyers, B.C.; Faridi, N.; Chin, D.B.; Stelly, D.M.; Mitchelmore, R.W.

submitted to the EMBL Data Library, August 1997

A:Reference number: Z16412

A:Accession: T08421

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1139 <SHE>  
 A:Cross-references: EMBL:AF017752; NID:g2852685; PID:g2852686  
 C:Genetics:  
 A:Gene: RGC2a  
 A:Introns: 892/3

Query Match 75.0%; Score 33; DB 2; Length 1139;  
 Best Local Similarity 55.6%; Pred. No. 1.6e+02;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 QYLRALGLK 9  
 ||| :|||  
 Db 31 QYMRMGK 39

RESULT 11

S31262  
 Tyb protein - yeast (*Saccharomyces cerevisiae*) retrotransposon Ty4 (fragment)  
 C:Species: *Saccharomyces cerevisiae*  
 C>Date: 18-Jun-1993 #sequence\_revision 25-Apr-1997 #text\_change 26-May-2000  
 C:Accession: S31262  
 R:Jancetky, B.; Lehle, L.  
 J. Biol. Chem. 267, 19798-19805, 1992  
 A:Title: Ty4, a new retrotransposon from *Saccharomyces cerevisiae*, flanked by tau-element  
 A:Reference number: S31261; MUID:93015829; PMID:1328182  
 A:Accession: S31262  
 A:Molecule type: DNA  
 A:Residues: 1-1465 <JAN>  
 A:Cross-references: EMBL:S46865  
 C:Genetics:  
 A:Mobile element: retrotransposon Ty4  
 C:Superfamily: Tyb protein

Query Match 75.0%; Score 33; DB 2; Length 1465;  
 Best Local Similarity 66.7%; Pred. No. 2.1e+02;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 QYLRALGLK 9  
 ||| :|||  
 Db 1094 QYLRIGL 1102

RESULT 12

PC1253  
 Tyb protein - yeast (*Saccharomyces cerevisiae*) retrotransposon Ty4 (fragment)  
 N:Contains: integrase; proteinase; reverse transcriptase; RNase H  
 C:Species: *Saccharomyces cerevisiae*  
 C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 26-May-2000  
 C:Accession: PC1253; S27465  
 R:Stucka, R.; Schwarzlöse, C.; Lochmüller, H.; Haacker, U.; Feldmann, H.  
 Gene 122, 119-128, 1992  
 A:Title: Molecular analysis of the yeast Ty4 element: Homology with Ty1, copia, and plant  
 A:Reference number: J1482; MUID:93083972; PMID:1333437  
 A:Accession: PC1253  
 A:Molecule type: DNA  
 A:Residues: 1-1467 <STU>  
 R:Stucka, R.; Schwarzlöse, C.; Lochmüller, H.; Haacker, U.; Feldmann, H.  
 submitted to the EMBL Data Library, May 1992  
 A:Description: Molecular analysis of the yeast Ty4 element: homology with Ty1, copia, an  
 A:Reference number: S27465  
 A:Accession: S27465  
 A:Molecule type: DNA  
 A:Residues: 1-1467 <STW>  
 A:Cross-references: EMBL:M94164  
 C:Genetics:  
 A:Gene: TY4B  
 A:Mobile element: retrotransposon Ty4  
 C:Superfamily: Tyb protein

Query Match 75.0%; Score 33; DB 2; Length 1467;  
 Best Local Similarity 66.7%; Pred. No. 2.1e+02;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 QYLRALGLK 9  
 ||| :|||  
 Db 1094 QYLRIGL 1102

RESULT 13

S52611  
 Tyb protein - yeast (*Saccharomyces cerevisiae*) retrotransposon Ty4  
 N:Alternate names: protein YHL008w-a  
 C:Species: *Saccharomyces cerevisiae*  
 C>Date: 05-May-1995 #sequence\_revision 19-Oct-1995 #text\_change 23-Mar-2001  
 C:Accession: S52611  
 R:Favell, T.  
 submitted to the EMBL Data Library, June 1994  
 A:Description: The sequence of S. *cerevisiae* cosmid L5018.  
 A:Reference number: S46798  
 A:Accession: S52611  
 A:Molecule type: DNA  
 A:Residues: 1-1802 <FAV>  
 A:Cross-references: EMBL:U11581  
 C:Genetics:  
 A:Map position: 8L  
 A:Mobile element: retrotransposon Ty4  
 C:Superfamily: Tyb protein

Query Match 75.0%; Score 33; DB 2; Length 1802;  
 Best Local Similarity 66.7%; Pred. No. 2.6e+02;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 QYLRALGLK 9  
 ||| :|||  
 Db 1431 QYLRIGL 1439

RESULT 14

S56894  
 Tyb protein - yeast (*Saccharomyces cerevisiae*) retrotransposon Ty4.JL  
 N:Alternate names: protein J0780; protein YJL113w  
 C:Species: *Saccharomyces cerevisiae*  
 C>Date: 08-Jul-1995 #sequence\_revision 08-Sep-1995 #text\_change 17-Mar-2000  
 C:Accession: S56894  
 R:Cziapluch, C.; Kordes, E.; Pujol, A.; Jauniaux, J.C.  
 submitted to the Protein Sequence Database, September 1995  
 A:Reference number: S56894  
 A:Accession: S56894  
 A:Molecule type: DNA  
 A:Residues: 1-1803 <CZJ>  
 A:Cross-references: EMBL:Z49389  
 C:Genetics:  
 A:Map position: 10L  
 A:Mobile element: retrotransposon Ty4.JL  
 C:Superfamily: Tyb protein

Query Match 75.0%; Score 33; DB 2; Length 1803;  
 Best Local Similarity 66.7%; Pred. No. 2.6e+02;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 QYLRALGLK 9  
 ||| :|||  
 Db 1432 QYLRIGL 1440

RESULT 15

E84232  
 hypothetical protein Vng0750c [imported] - *Halobacterium* sp. NRC-1  
 C:Species: *Halobacterium* sp. NRC-1  
 C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C:Accession: E84232  
 R:Ng, W.V.; Kennedy, S.P.; Mahanas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky  
 ; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja  
 Jung, K.H.; Alm, M.; Freitas, T.  
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
 A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.;

A:Title: Genome sequence of Halobacterium species NRC-1.  
A:Reference number: A84160; MUID:20504483; PMID:11016950  
A:Accession: E84232  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-239 <STO>  
A:Cross-references: GB:AE004437; NID:g10580329; PIDN:AAG19225.1; GSPDB:GN00138  
C:Genetics:  
A:Gene: VNG0750C

Query Match 72.7%; Score 32; DB 2; Length 239;  
Best Local Similarly 85.7%; Pred. No. 54;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Caps 0;

OY 2 YLRALGL 8  
| | | | | | | |  
Db 6 YLRALGL 12

Search completed: July 11, 2003, 12:00:38  
Job time : 3.91497 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:43:41 : Search time 0.860707 Seconds

(Without alignments)  
433.698 Million cell updates/sec

Title: US-10-091-442-5

Sequence: 1 QYLRALGLK 9

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44	100.0	380	1 SPB7_HUMAN	O75635 homo sapien
2	35	79.5	164	1 PPOC_BACSU	P39789 bacillus su
3	34	77.3	350	1 ADH1_CANAL	P43067 candida alb
4	33	75.0	127	1 YEG1_MERJA	O58886 methanococc
5	33	75.0	1803	1 YJL3_YEAST	P47024 saccharomyc
6	32	72.7	434	1 Y610_MERJA	O58021 methanococc
7	32	72.7	487	1 YFGC_SALTY	O68975 salmonella
8	32	72.7	654	1 ICAL_RAT	P27321 rattus norv
9	32	72.7	788	1 ICAL_MOUSE	P51125 mus musculu
10	31	70.5	349	1 RECA_CLOAB	P58254 clostridium
11	31	70.5	352	1 ADH3_EMENI	P07754 emeritella
12	31	70.5	355	1 ISPG_THETN	Q8ra30 thermoaer
13	31	70.5	383	1 HIS2_AOUAE	O67223 aquifex aeo
14	31	70.5	403	1 DACA_ECOLI	P04287 escherichia
15	31	70.5	619	1 GIDA_CAMJE	O99na7 campylobact
16	31	70.5	670	1 DNLJ_BACST	O67703 bacillus st
17	31	70.5	960	1 CHPL_SCHRO	Q10103 schizosacch
18	30	68.2	262	1 V243_FOWPV	O91427 fowlpox vir
19	30	68.2	325	1 NARI_MOUSE	O60935 mus musculu
20	30	68.2	327	1 NARI_HUMAN	P53961 homo sapien
21	30	68.2	331	1 GLK_CAUCR	O9a6n3 caulobacter
22	30	68.2	336	1 XYLZ_PSEPU	P27101 pseudomonas
23	30	68.2	345	1 H10W_BOVIN	P10950 bos taurus
24	30	68.2	375	1 ADH4_KLULA	P49385 kluyveromyc
25	30	68.2	387	1 DHAT_CITFR	P45513 citrobacter
26	30	68.2	387	1 ISCS_BUCAP	O59477 klebsiella
27	30	68.2	404	1 ISCS_BUCAP	O59477 klebsiella
28	30	68.2	419	1 LMB2_YERPE	O8rhp0 yersinia pe
29	30	68.2	449	1 MURC_HELPJ	O9a112 helicobacte
30	30	68.2	449	1 MURC_HELPJ	O9a112 helicobacte
31	30	68.2	470	1 ROS2_HELPY	O62340 helicobacte
32	30	68.2	992	1 PRO_LEPAV	O01115 leptosphaer
33	30	68.2	1327	1 TWK1_HUMAN	O95271 homo sapien

34	30	68.2	1363	1 XDH_EMENI	Q12553 emeritella
35	30	68.2	1473	1 OVOS_CHICK	P20740 gallus gall
36	30	68.2	1545	1 MRP2_HUMAN	O92887 homo sapien
37	30	68.2	1666	1 MYM1_MOUSE	O62233 mus musculu
38	30	68.2	2051	1 FAS1_YEAST	P07149 s fatly aci
39	29	65.9	60	1 RL30_STRCO	P46789 streptomyc
40	29	65.9	74	1 VSH2_MICIU	P33104 micrococcus
41	29	65.9	141	1 VSH2_HAEPA	P36434 haemophilus
42	29	65.9	162	1 PHZB_PSEFL	O51788 pseudomonas
43	29	65.9	170	1 YBET_ECOLI	O57822 methanococc
44	29	65.9	184	1 YBET_ECOLI	P77296 escherichia
45	29	65.9	190	1 Y516_AERPE	O9yer5 aeropyrum p

## ALIGNMENTS

RESULT 1  
ID SPB7\_HUMAN STANDARD: PRT: 380 AA.  
AC O75635  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Megsin (TP55) (Serpin B7).  
GN SERPINB7.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Carnivora; Hominoidea; Homo.  
OX NCBI\_TaxID:9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97326116; PubMed=9182567;  
RA Tsurumoto M., Tsurumoto N., Ishida N., Kurihara T., Iwasa F.,  
RA Yamashiro K., Rogi T., Kodama S., Katsunagi N., Adachi M.,  
RA Katayama T., Nakao M., Yamauchi K., Hashino J., Haruyama M., Miura K.,  
RA Nakanishi T., Nakazato H., Teramura M., Mitsuuchi H., Yamaguchi N.,  
RT "Purification, CDNA cloning, and characterization of a new serpin with  
RT megakaryocyte maturation activity.",  
RT J. Biol. Chem. 272:15373-15380(1997).  
RL (2)  
RN SEQUENCE FROM N.A.  
RP TISSUE-Mesangial cells;  
RX MEDLINE=98376492; PubMed=9710452;  
RA Miyata T., Nangaku M., Suzuki D., Inagi R., Uragami K., Sakai H.,  
RA Okubo K., Kurokawa K.,  
RT "A mesangium-predominant gene, megasin, is a new serpin upregulated in  
RT J. Clin. Invest. 102:828-836(1998).  
RL  
CC -!- FUNCTION: Might function as an inhibitor of lys-specific  
CC proteases. Might influence the maturation of megakaryocytes via  
CC its action as a serpin.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -!- TISSUE SPECIFICITY: Predominantly expressed in mesangial cells.  
CC -!- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
-----  
CC EMBL: D88575; BAA3123.1; -  
CC EMBL: AF027866; AAC64506.1; -  
CC HSSP: P05619; IHLE.  
CC Genew: HGNC:13902; SERPINB7.  
CC MIM: 603357; -  
CC InterPro: IPR000215; Serpin.  
CC Pfam: PF00079; serpin; 1.  
CC SMART: SM00093; SERPIN; 1.  
CC PROSITE: PS00284; SERPIN; 1.

KW Serpin: Serine protease inhibitor.  
 FT ACT SITE 347 348 REACTIVE BOND (BY SIMILARITY).  
 SQ SEQUENCE 380 AA; 42904 MW; 9A2CDB6C63CFF605 CRC64;

Query Match 100.0%; Score 44; DB 1; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 0.11;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYRALGK 9  
 DB 289 QYRALGK 297

RESULT 2  
 YPOC\_BACSU  
 ID YPOC\_BACSU STANDARD; PRT; 164 AA.  
 AC P39789;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein ypoC (ORF).  
 GN YPOC OR YPOC.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=95219085; PubMed=7704260;  
 RA Bruand C., Sorokin A., Serror P., Ehrlich S.D.;  
 RT "Nucleotide sequence of the Bacillus subtilis dnaD gene."  
 RL Microbiology 141:321-322(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=95113769; PubMed=7814321;  
 RA Popham D.L., Setlow P.;  
 RT "Cloning, nucleotide sequence, and mutagenesis of the Bacillus subtilis pona operon, which codes for penicillin-binding protein (PBP) 1 and a PBP-related factor."  
 RL J. Bacteriol. 177:326-335(1995).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kuntz F., Ogatawara N., Moszer I., Albertini A.M., Alloni G.,  
 AZEVEDO V., Betero M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Bortiss R., Bourcier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Conerton I.F., Cummings S.D., Daniel R.A.,  
 RA Dentzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Enlian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Funa S., Galizzi A., Galleron N.,  
 RA Ghim S.Y., Glaser P., Goffeau A., Golligly E.J., Grandi G.,  
 RA Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holseppel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Bianchard M., Klein C.,  
 RA Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kunano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Neome D., O'Reilly M., Ogawa K., Ogilwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Potworlik S., Prescott A.M.,  
 RA Pirescen E., Pujic P., Purnelle B., Rapoport G., Ray M., Reynolds S.,  
 RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadate Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Serror S.J., Serror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni K.,  
 RA Toso V., Uchiyama S., Vandenbol M., Vannier F., Vassaret A.,  
 RA Vari A., Wamutit R., Medler E., Medler H., Meitzenger T.,  
 RA Winers P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,

RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
 RT subtilis."  
 RL Nature 390:249-256(1997).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: U11289; AAA80006.1; -  
 DR EMBL: U11883; AAA64948.1; -  
 DR EMBL: L47709; AAB38458.1; -  
 DR EMBL: Z99115; CAB14149.1; -  
 DR Sublist: BG10955; YPOC.  
 KW Hypothetical protein: Complete proteome.  
 SQ SEQUENCE 164 AA; 19346 MW; 5BB082046B7674CB CRC64;

Query Match 79.5%; Score 35; DB 1; Length 164;  
 Best Local Similarity 77.8%; Pred. No. 3.5;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 QYRALGK 9  
 DB 13 QYRALGK 21

RESULT 3  
 ADH1\_CANAL  
 ID ADH1\_CANAL STANDARD; PRT; 350 AA.  
 AC P43067;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Alcohol dehydrogenase I (Ec 1.1.1.1) (40 kDa allergen) (Allergen  
 DE can a 1) (can a 1).  
 GN ADH1 OR CAD.  
 OS Candida albicans (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
 OX NCBI\_TaxID=5476;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96287648; PubMed=8686375;  
 RA Bertam G., Swoboda R.K., Gooday G.W., Gow N.A.R., Brown A.J.P.;  
 RT "Structure and regulation of the Candida albicans ADH1 gene encoding  
 RT an immunogenic alcohol dehydrogenase."  
 RL Yeast 12:115-127(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B311A;  
 RA Pendrak M.L., Klotz S.A., Smith R.L.;  
 RL Submitted (Oct-1994) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 104-313 FROM N.A.  
 RX MEDLINE=92136159; PubMed=1777830;  
 RA Shen H.D., Choo K.B., Lee H.H., Hsieh J.C., Lin W.L., Lee W.R.,  
 RA Han S.H.;  
 RT "The 40-kilodalton allergen of Candida albicans is an alcohol  
 RT dehydrogenase: molecular cloning and immunological analysis using  
 RT monoclonal antibodies."  
 RL Clin. Exp. Allergy 21:675-681(1991).  
 CC -1- CATALYTIC ACTIVITY: An alcohol + NAD(+) = an aldehyde or ketone +  
 CC NADH.  
 CC -1- COFACTOR: ZINC.  
 CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
 CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE  
 CC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration



CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X81694; CAA57342.1; -  
 DR EMBL: U15924; AAA53300.1; -  
 DR COMPLEYEAST-2DPAGE; P43067; -  
 DR InterPro: IPR002328; Adh\_zinc.  
 DR Pfam: PF00107; adh\_zinc; 1.  
 DR PROSITE: PS00059; Adh\_zinc; 1.  
 KM Oxidoreductase; Zinc; NAD; Allergen.  
 FT METAL 46 46 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 69 69 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 100 100 ZINC (SECOND ATOM) (BY SIMILARITY).  
 FT METAL 103 103 ZINC (SECOND ATOM) (BY SIMILARITY).  
 FT METAL 106 106 ZINC (SECOND ATOM) (BY SIMILARITY).  
 FT METAL 114 114 ZINC (SECOND ATOM) (BY SIMILARITY).  
 FT METAL 156 156 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT CONFLICT 39 39 H -> N (IN REF. 2).  
 FT CONFLICT 53 53 R -> W (IN REF. 2).  
 FT CONFLICT 140 140 F -> T (IN REF. 3).  
 FT CONFLICT 212 212 F -> L (IN REF. 2).  
 FT CONFLICT 227 227 D -> A (IN REF. 3).  
 FT CONFLICT 313 313 R -> S (IN REF. 3).  
 FT CONFLICT 327 327 D -> E (IN REF. 2).  
 SQ SEQUENCE 350 AA; 36879 MW; C3DFE2E70F42D634 CRC64;

Query Match 77.3%; Score 34; DB 1; Length 350;  
 Best Local Similarity 66.7%; Pred. No. 12;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 QYRALGLK 9  
 DB 191 QYRAMGLR 199

RESULT 4  
 ID YE91\_METJA STANDARD; PRT; 127 AA.  
 AC Q58886;  
 DT 16-OCT-2001 (Rel. 40; Last sequence update)  
 DT 16-OCT-2001 (Rel. 40; Last annotation update)  
 DE Hypothetical protein MJ1491.  
 GN MJ1491.  
 OS Methanococcus jannaschii.  
 OC Archaea: Euryarchaeota: Methanococci: Methanococcales;  
 CC Methanocaldococcaceae: Methanocaldococcus.  
 OX NCBI\_TaxID=2190;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE=96337999; PubMed=8688087;  
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
 RA Overbeek R., Kirkness E.F., Weisscock K.G., Merrick J.M., Glodek A.,  
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fummann J.L., Nguyen D.,  
 RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hursb M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 RT jannaschii.";  
 RT Science 273:1058-1073(1996).  
 RL -1- SUBCELLULAR LOCATION: Integral membrane protein (potential)  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: U67590; AAB99510.1; -  
 DR TIGR: MJ1491; -  
 KM Hypothetical protein: Transmembrane; Complete proteome.  
 FT TRANSMEM 16 36 POTENTIAL.  
 FT TRANSMEM 59 79 POTENTIAL.  
 FT TRANSMEM 100 120 POTENTIAL.  
 SQ SEQUENCE 127 AA; 13006 MW; 6D3FB594A7961C96 CRC64;

Query Match 75.0%; Score 33; DB 1; Length 127;  
 Best Local Similarity 87.5%; Pred. No. 7.1;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 YLRALGLK 9  
 DB 83 YLSALGLK 90

RESULT 5  
 ID YJL3\_YEAST STANDARD; PRT; 1803 AA.  
 AC P47024; P87192;  
 DT 01-FEB-1996 (Rel. 33; Created)  
 DT 30-MAY-2000 (Rel. 39; Last sequence update)  
 DT 30-MAY-2000 (Rel. 39; Last annotation update)  
 DE Transposon Ty4 207.7 kDa hypothetical protein.  
 GN TY4B OR YJL113W OR J0780.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S288C / FY1679;  
 RX MEDLINE=97103775; PubMed=8948101;  
 RA Cziepluch C., Kordes E., Pujol A., Jauniaux J.-C.;  
 RT "Sequencing analysis of a 40.2 kb fragment of yeast chromosome X  
 RT reveals 19 open reading frames including URA2 (5' end), TRK1, PBS2,  
 RT SPRI0, GC014, Rpt1, PHO86, NCA3, ASF1, CCT7, GZF3, two tRNA genes,  
 RT three remnant delta elements and a Ty4 transposon.";  
 RL Yeast 12:1471-1474(1996).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: 249389; CAA89409.1; -  
 DR SCD: S0003649; YJL113W.  
 DR InterPro: IPR001584; Rve.  
 DR InterPro: IPR001878; Znf\_CCHC.  
 DR Pfam: PF00665; rve; 1.  
 DR SMART: SM00343; ZNF\_C2HC; 1.  
 KW Transposable element; Hypothetical protein.  
 SQ SEQUENCE 1803 AA; 207693 MW; 16DCD7284ABD52D3 CRC64;

Query Match 75.0%; Score 33; DB 1; Length 1803;  
 Best Local Similarity 66.7%; Pred. No. 1e+02;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 QYRALGLK 9  
 DB 1432 QYLCIGLK 1440

RESULT 6

```

Y610_METUA
ID 1610_METUA STANDARD; PRT: 434 AA.
AC Q58027;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein M0610.
GN M0610.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96371999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodok A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
RA Uitterlinden T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kalne B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT *Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.;
RL Science 273:1058-1073(1996).
CC -1- SIMILARITY: TO C.ELEGANS D1054.13.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: U67509; AAB96603.1; -.
CC DR TIGR: M00610; -.
CC KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 434 AA; 48683 MW; 09AFBC5819613EF5 CRC64;

Query Match 72.7%; Score 32; DB 1; Length 434;
Best Local Similarity 75.0%; Pred. No. 39;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 YLRALGLK 9
Db 121 FLNALGLK 128

RESULT 7
YFGC_SALTY STANDARD; PRT: 487 AA.
ID YFGC_SALTY STANDARD; PRT: 487 AA.
AC Q8XG75;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein yfgc precursor.
DE YFGC OR STM2494 OR STY2735.
GN YFGC OR STM2494 OR STY2735.
OS Salmonella typhimurium, and
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602, 601;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S. typhimurium; STRAIN=LT2 / SGC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McEllelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Potvin L.S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,

```

```

RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S. typhi; STRAIN=CR18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks R., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagsels K.,
RA Krogan A., Larsen T.S., Leather S., Moule S., O'Gara P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CR18.";
RL Nature 413:848-852(2001).
CC -1- SIMILARITY: CONTAINS 2 TPR REPEATS.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: AE008812; AL21388.1; -.
CC DR STYGENE; S677777; YFGC.
DR StyGene; S677777; YFGC.
KW InterPro: IPR001440; TPR.
RT Hypothetical protein; Signal; Repeat; TPR repeat; Complete proteome.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 487 HYPOTHETICAL PROTEIN YFGC.
FT REPEAT 309 342 TPR 1.
FT REPEAT 427 460 TPR 2.
SQ SEQUENCE 487 AA; 53740 MW; 593FCFC8DBCCCE CRC64;

Query Match 72.7%; Score 32; DB 1; Length 487;
Best Local Similarity 55.6%; Pred. No. 44;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 QYRALGLK 9
Db 74 QYINALGLK 82

RESULT 8
ICAL_RAT STANDARD; PRT: 654 AA.
ID ICAL_RAT STANDARD; PRT: 654 AA.
AC P27321;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Calpain inhibitor (Calpastatin).
DE CAST.
GN Rattus norvegicus (Rat).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=91198152; PubMed=2015306;
RA Ishida S., Emori Y., Suzuki K.;
RT "Rat calpastatin has diverged primary sequence from other mammalian
RT calpastatins but retains functionally important sequences.";
RL Biochim. Biophys. Acta 1088:436-438(1991).
RN [2]
RP SEQUENCE OF 1-129 FROM N.A., AND ALTERNATIVE SPLICING.
RC STRAIN=Fischer;

```

RX MEDLINE-92110395; PubMed-1730065;  
 RA Lee W.J., Hatanaka M., Maki M.;  
 RT "Multiple forms of rat calpastatin cDNA in the coding region of  
 functionally unknown amino-terminal domain.";  
 RL Biochim. Biophys. Acta 1129:251-253(1992).  
 CC -1- FUNCTION: Specific inhibition of calpain (calcium-dependent  
 cysteine protease). Plays a key role in postmortem tenderization  
 of meat and have been proposed to be involved in muscle protein  
 degradation in living tissue.  
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS: A (SHOWN HERE), B AND C; ARE  
 PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- DOMAIN: HAS FOUR INHIBITORY DOMAINS.  
 CC -1- SIMILARITY: BELONGS TO THE CALPASTATIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X56729; CAA40053.1; -  
 DR EMBL: X62520; CAA44386.1; -  
 DR PIR: S15074; S15074.  
 DR InterPro: IPR001259; Calpain\_inhnb.  
 DR Pfam: PF00748; Calpain\_inhnb: 4.  
 KM Repeat: Thiol protease inhibitor: Alternative splicing.  
 FT REPEAT 149 201 INHIBITORY DOMAIN 1.  
 FT REPEAT 282 334 INHIBITORY DOMAIN 2.  
 FT REPEAT 392 445 INHIBITORY DOMAIN 3.  
 FT REPEAT 529 582 INHIBITORY DOMAIN 4.  
 FT VARSPPLIC 8 20 MISSING (IN ISOFORM C).  
 FT VARSPPLIC 39 76 MISSING (IN ISOFORM B AND ISOFORM C).  
 FT CONFLICT 102 102 K -> E (IN REF. 1).  
 SQ SEQUENCE 654 AA; 71367 MW; 4970B9A198D819A CRC64;  
 Query Match 72.7%; Score 32; DB 1; Length 654;  
 Best Local Similarity 75.0%; Pred. No. 59;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 YLRALGLK 9  
 Db 169 YLEALGIK 176  
 RESULT 9  
 ID ICAL\_MOUSE STANDARD; PRT; 788 AA.  
 AC P51125; O9RON1; O9OX03; O9OX04.  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 15-JUN-2003 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Calpain inhibitor (Calpastatin).  
 DE CAST.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID:10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS MCS-A; MCS-B AND MCS-C).  
 RC STRAIN=C3H; TISSUE=Muscle;  
 RX MEDLINE-99333681; PubMed-10403772;  
 RA Takano J., Kawamura T., Murase M., Hltomi K., Maki M.;  
 RT "Structure of mouse calpastatin isoforms: implications of species-  
 common and species-specific alternative splicing.";  
 RL Biochem. Biophys. Res. Commun. 260:339-345(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORMS TCAST1 AND TCAST2).  
 RC STRAIN=CD-1;  
 RA Li S., Goldberg E.;  
 RT "Characterization of a membrane associated testis-specific  
 calpastatin.";

RL Submitted (SEP-1999) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 84-363 FROM N.A. (ISOFORM MCS-B).  
 RC STRAIN=BA1B/C; TISSUE=Heart;  
 RX MEDLINE-92110395; PubMed-1730065;  
 RA Lee W.J., Hatanaka M., Maki M.;  
 RT "Multiple forms of rat calpastatin cDNA in the coding region of  
 functionally unknown amino-terminal domain.";  
 RL Biochim. Biophys. Acta 1129:251-253(1992).  
 CC -1- FUNCTION: Specific inhibition of calpain (calcium-dependent  
 cysteine protease). Plays a key role in postmortem tenderization  
 of meat and have been proposed to be involved in muscle protein  
 degradation in living tissue.  
 CC -1- DOMAIN: HAS FOUR INHIBITORY DOMAINS.  
 CC -1- ALTERNATIVE PRODUCTS: 5 ISOFORMS: MCS-A (SHOWN HERE), MCS-B, MCS-  
 C, TCAST1 AND TCAST2; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: ISOFORM MCS-B IS THE MAJOR FORM IN ALL TISSUES  
 EXAMINED. ISOFORM MCS-A ACCOUNTS FOR 5-10% IN TISSUES SUCH AS  
 SKELETAL MUSCLE, LIVER AND BRAIN, AND 30% IN MYOBLASTS. ISOFORMS  
 TCAST1 AND TCAST2 ARE TESTIS-SPECIFIC.  
 CC -1- SIMILARITY: BELONGS TO THE CALPASTATIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AB026997; BAA84768.1; -  
 DR EMBL: AF190152; AAF25194.1; -  
 DR EMBL: AF190151; AAF25193.1; -  
 DR EMBL: X62519; CAA44385.1; -  
 DR MGD: MGI:1098236; Cast.  
 DR InterPro: IPR001259; Calpain\_inhnb.  
 DR Pfam: PF00748; Calpain\_inhnb: 2.  
 KM Repeat: Thiol protease inhibitor: Alternative splicing.  
 FT REPEAT 251 303 INHIBITORY DOMAIN 1.  
 FT REPEAT 384 436 INHIBITORY DOMAIN 2.  
 FT REPEAT 524 577 INHIBITORY DOMAIN 3.  
 FT REPEAT 661 714 INHIBITORY DOMAIN 4.  
 FT VARSPPLIC 1 341 MISSING (IN ISOFORM TCAST1).  
 FT VARSPPLIC 1 357 MISSING (IN ISOFORM TCAST2).  
 FT VARSPPLIC 92 110 MISSING (IN ISOFORM MCS-B AND ISOFORM  
 MCS-C).  
 FT VARSPPLIC 305 333 MISSING (IN ISOFORM MCS-C).  
 FT VARSPPLIC 342 396 PDPSVHSQAEOVK -> MGQFLSTFWGSPAAVWOEKLK  
 EGERKAGETIPILDDHYVCSEREHSGKH (IN  
 ISOFORM TCAST1).  
 FT KRVKEEVINDALALSLSTGTRPPDPSPHSQAEOYKE  
 -> MGQFLSTFWGSPAAVWOEKLKREGERKAGETIPILQ  
 DH (IN ISOFORM TCAST2).  
 SQ SEQUENCE 788 AA; 84921 MW; 28E4D3B4A68BEB9 CRC64;  
 Query Match 72.7%; Score 32; DB 1; Length 788;  
 Best Local Similarity 75.0%; Pred. No. 71;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 YLRALGLK 9  
 Db 271 YLEALGIK 278  
 RESULT 10  
 ID RECA\_C10AB STANDARD; PRT; 349 AA.  
 AC P58254;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Recl protein (Recombinase A).

GN RECA OR CAC1815.  
 OS Clostridium acetobutylicum.  
 CC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 CC Clostridium.  
 CC NCBI\_TaxID=1488;  
 CC [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 824 / DSM 792 / VKM B-1787;  
 RX MEDLINE-21359325; Pubmed-11466286;  
 RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,  
 Gibson R.L., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,  
 Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,  
 Bennett G.N., Koonin E.V., Smith D.R.;  
 RT Genome sequence and comparative analysis of the solvent-producing  
 bacterium Clostridium acetobutylicum.";  
 RL J. Bacteriol. 185:4823-4838(2001).  
 CC -1- FUNCTION: CAN CATALYZE THE HYDROLYSIS OF ATP IN THE PRESENCE OF  
 CC SINGLE-STRANDED DNA. THE ATP-DEPENDENT UPTAKE OF SINGLE-STRANDED  
 CC DNA BY DUPLEX DNA, AND THE ATP-DEPENDENT HYBRIDIZATION OF  
 CC HOMOLOGOUS SINGLE-STRANDED DNAs. IT INTERACTS WITH LEXA CAUSING  
 CC ITS ACTIVATION AND LEADING TO ITS AUTOCATALYTIC CLEAVAGE (BY  
 CC SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE RECA FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: AE007690; AAK9780.1; -  
 DR InterPro: IPR003593; AAA\_ATPase.  
 DR InterPro: IPR001553; RECA.  
 DR Pfam: PF00154; RECA.1.  
 DR PRINTS: PR00142; RECA.  
 DR PRODOM: PD000229; RECA.1.  
 DR SMART: SM00382; AAA.1.  
 DR PROSITE: PS00321; RECA.1; 1.  
 DR PROSITE: PS50162; RECA.2; 1.  
 DR PROSITE: PS50163; RECA.3; 1.  
 KM DNA damage: DNA recombination; SOS response; ATP-binding; DNA-binding;  
 KM Complete proteome.  
 FT NP\_BIND 65 72 ATP (By similarity).  
 FT SEQUENCE 349 AA; 38089 MW; ICB821EDFB3C855D CRC64;  
 SQ  
 Query Match 70.5%; Score 31; DB 1; Length 349;  
 Best Local Similarity 75.0%; Pred. No. 51;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 QYRALGL 8  
 II :|||:  
 DB 101 QYRALGLV 108  
 RESULT 11  
 ADH3\_EMENT STANDARD; PRT; 352 AA.  
 ID ADH3\_EMENT  
 AC P07754;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-AUG-1988 (Rel. 08, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Alcohol dehydrogenase III (EC 1.1.1.1) (ADH III).  
 GN ALDC OR ADH3.  
 OS *Escherichia nidulans* (Aspergillus nidulans).  
 CC Eukaryotes; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 CC Eurotiales; Trichocomaceae; *Escherichia*.  
 CC NCBI\_TaxID=5072;  
 CC [1]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=86055769; Pubmed=2998782;

RA McKnight L., Kato H., Upshall A., Parker M.D., O'Hara P.J., O'Hara S.;  
 RT Identification and molecular analysis of a third *Aspergillus*  
 RT *nidulans* alcohol dehydrogenase gene.";  
 RL EMO J. 4:2093-2099(1985).  
 CC -1- CATALYTIC ACTIVITY: An alcohol + NAD(+) = an aldehyde or ketone +  
 CC NADH.  
 CC -1- COFACTOR: ZINC.  
 CC -1- SUBUNIT: HOMOTETRAMER.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE  
 CC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: X02764; CAA26541.1; -  
 DR PIR: A24648; A24648.  
 DR InterPro: IPR002328; Adh\_zinc.  
 DR InterPro: IPR002085; Adh\_zn\_family.  
 DR Pfam: PF00107; adh\_zinc.1.  
 DR PROSITE: PS00059; ADH\_ZINC; 1.  
 KM Oxidoreductase; zinc; NAD; Multigene family.  
 FT METAL 44 44 ZINC (Catalytic).  
 FT METAL 67 67 ZINC (Catalytic).  
 FT METAL 100 100 ZINC (SECOND ATOM).  
 FT METAL 103 103 ZINC (SECOND ATOM).  
 FT METAL 106 106 ZINC (SECOND ATOM).  
 FT METAL 114 114 ZINC (SECOND ATOM).  
 FT METAL 156 156 ZINC (Catalytic).  
 FT SEQUENCE 352 AA; 37126 MW; AFDE78FE442E3144 CRC64;  
 SQ  
 Query Match 70.5%; Score 31; DB 1; Length 352;  
 Best Local Similarity 55.6%; Pred. No. 51;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 QYRALGL 9  
 II :|||:  
 DB 191 QYRALGLR 199  
 RESULT 12  
 ISPG\_THETN STANDARD; PRT; 355 AA.  
 ID ISPG\_THETN  
 AC OBR430;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.  
 GN ISPG OR GCPE OR TREL400.  
 OS *Thermoplasma acidophilum* (Thermoplasma).  
 CC Bacteria; Firmicutes; Clostridia; Thermoplasmales;  
 CC Thermoplasmales; Thermoplasmales; Thermoplasmales;  
 CC NCBI\_TaxID=119072;  
 CC [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-M84 / JCM 11007;  
 RX MEDLINE-21992816; Pubmed-11997336;  
 RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,  
 RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,  
 RA Tan H., Chen R., Wang J., Yu J., Yang H.;  
 RT "A complete sequence of T. tengcongensis genome.";  
 RL Genome Res. 12:689-700(2002).  
 CC -1- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate  
 CC (ME-2,4CP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate  
 CC (By similarity).  
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; sixth step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPG FAMILY.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AE013100; AAM24622.1;  
 KW Isoprene biosynthesis: Complete proteome.  
 SQ SEQUENCE 355 AA; 39005 MW; ADDEP97DD5622A0 CRC64;

Query Match 70.5%; Score 31; DB 1; Length 355;  
 Best Local Similarity 87.5%; Pred. No. 52;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 QYLRALGL 8  
 Db 249 QILRALGL 256

RESULT 13  
 HIS2\_AOUAE STANDARD; PRT; 383 AA.  
 ID HIS2\_AOUAE  
 AC O67223;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE ATP phosphoribosyltransferase regulatory subunit.  
 GN HIS2 OR HIS52 OR AQ\_1155.  
 OS Aquifex aeolicus.  
 OC Bacteria; Aquificae; Aquificaceae; Aquificaceae;  
 OX Aquifex.  
 RX NCBI\_TaxID=63363;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VF5;  
 RA MEDLINE=98196666; PubMed=9537320;  
 RA Decker G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
 RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Anjay M., Huber R.,  
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;  
 RA "The complete genome of the hyperthermophilic bacterium Aquifex  
 RA aeolicus";  
 RT Nature 392:353-358(1998).

RL "The complete genome of the hyperthermophilic bacterium Aquifex  
 CC -1- FUNCTION: May allow the regulation of ATP  
 CC phosphoribosyltransferase activity by histidine (By similarity).  
 CC -1- PATHWAY: Histidine biosynthesis; first step. Very important in the  
 CC regulation of histidine metabolism.  
 CC -1- SUBUNIT: Binds to hisG (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
 CC -1- MISCELLANEOUS: This function is generally fulfilled by the C-  
 CC terminal part of hisG, which is missing in some bacteria such as  
 CC this one.  
 CC -1- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.  
 CC HIS2 SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AE000726; AAC07186.1;  
 DR HSSP: P04804; IKM.  
 DR InterPro: IPR002106; AATRNA\_1lgaseII.  
 DR InterPro: IPR004517; His2\_second.  
 DR InterPro: IPR002314; tRNA-synt\_2b.  
 DR Pfam: PF00587; tRNA-synt\_2b; 1.  
 DR TIGRFAMs: TIGR00443; his2\_second; 1.  
 DR Histidine biosynthesis: Complete proteome.  
 KW SEQUENCE 383 AA; 44349 MW; E3B449B5BEF86C8D CRC64;

Query Match 70.5%; Score 31; DB 1; Length 383;  
 Best Local Similarity 55.6%; Pred. No. 56;  
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Oy 1 QYLRALGLK 9  
 Db 140 EFLKSLGLK 148

RESULT 14

DNAC\_ECOLI STANDARD; PRT; 403 AA.  
 ID DNAC\_ECOLI  
 AC P04287; P77106;  
 DT 20-MAR-1987 (Rel. 04, Created)  
 DT 20-MAR-1987 (Rel. 04, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Penicillin-binding protein 5 precursor (D-alanyl-D-alanine  
 DE carboxypeptidase fraction A) (EC 3.4.16.4) (DD-peptidase)  
 DE (DD-carboxypeptidase) (PBP-5).  
 GN DACC OR PFV OR B0632 OR 20777 OR ECS0670.  
 OS Escherichia coli, and  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OX Escherichia.  
 RX NCBI\_TaxID=562, 83334;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RA MEDLINE=88157719; PubMed=3279397;  
 RA Broome-Smith J.K., Ioannidis I., Edelman A., Spratt B.G.;  
 RA "Nucleotide sequences of the penicillin-binding protein 5 and 6 genes  
 RA of Escherichia coli";  
 RT Nucleic Acids Res. 16:1617-1617(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Broome-Smith J.K.;  
 RL Submitted (Apr-1984) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RA MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12";  
 RL Science 277:1453-1474(1997).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RA MEDLINE=97061202; PubMed=8905232;  
 RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,  
 RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
 RA Yanai M., Horiiuchi T.;  
 RT "A 718-Kb DNA sequence of the Escherichia coli K-12 genome  
 RT corresponding to the 12.7-28.0 min region on the linkage map";  
 RL DNA Res. 3:137-155(1996).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,  
 RA Pederspiel N., Hyman R., Kaiman S., Komp C., Kurdi O., Lew H.,  
 RA Lin D., Mamath A., Oefner P., Roberts D., Schramm S., Davis R.W.;  
 RL Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
 RA MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,

RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
RA Grobbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoukis K.,  
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
RA Welch R.A., Blattner F.R.,  
RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.",  
RL Nature 409:529-533(2001).  
[7]  
RP SEQUENCE FROM N.A.  
RC STRAIN-O157:H7 / RIMD 0509952:  
RX MEDLINE-2155231; Pubmed-11238796;  
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,  
RA Kihara S., Shiba T., Hattori M., Shinagawa H.,  
RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*  
RT O157:H7 and genomic comparison with a laboratory strain K-12.",  
RL DNA Res. 8:11-22(2001).  
[8]  
RP SEQUENCE OF 108-336 FROM N.A., AND MUTANT DACA11191.  
RX MEDLINE-84108878; Pubmed-6319180;  
RA Broome-Smith J.K., Spratt B.G.,  
RT "An amino acid substitution that blocks the deacylation step in the  
RT enzyme mechanism of penicillin-binding protein 5 of *Escherichia*  
RT coli.",  
RL FEBS Lett. 165:185-189(1984).  
[9]  
RP SEQUENCE OF 1-39 FROM N.A.  
RC STRAIN-K12;  
RX MEDLINE-86058785; Pubmed-7042389;  
RA Takase I., Ishino F., Wachi M., Kamata H., Doi M., Asoh S.,  
RA Matsuzawa H., Ohta T., Matsushashi M.,  
RT "Genes encoding two lipoproteins in the *leuS*-*dacA* region of the  
RT *Escherichia coli* chromosome.",  
RL J. Bacteriol. 169:5692-5699(1987).  
[10]  
RP SEQUENCE OF 30-57.  
RX MEDLINE-82187192; Pubmed-7042389;  
RA Maxam D.J., Amanuma H., Strominger J.L.,  
RT "Amino acid sequence homologues between *Escherichia coli* penicillin-  
RT binding protein 5 and class A beta-lactamases.",  
RL FEBS Lett. 139:159-163(1982).  
[11]  
RP MUTAGENESIS OF LYS-213.  
RX MEDLINE-92283850; Pubmed-1597468;  
RA Malhotra K.T., Nicholas R.A.,  
RT "Substitution of lysine 213 with arginine in penicillin-binding  
RT protein 5 of *Escherichia coli* abolishes D-alanine carboxypeptidase  
RT activity without affecting penicillin binding.",  
RL J. Biol. Chem. 267:11386-11391(1992).  
[12]  
RP DOMAINS.  
RX MEDLINE-93143736; Pubmed-8424800;  
RA van der Linden M.P.G., de Haan L., Keck W.,  
RT "Domain organization of penicillin-binding protein 5 from *Escherichia*  
RT coli analysed by C-terminal truncation.",  
RL Biochem. J. 289:593-598(1993).  
[13]  
RP FUNCTION: REMOVES C-TERMINAL D-ALANYL RESIDUES FROM SUGAR-PEPTIDE  
CC CELL WALL PRECURSORS.  
CC -1- CATALYTIC ACTIVITY: D-alanyl-D-alanine + H(2)O = 2 D-alanine.  
CC -1- PATHWAY: FINAL STAGES IN PEPTIDOGLYCAN SYNTHESIS.  
CC -1- SUBCELLULAR LOCATION: INNER MEMBRANE-ASSOCIATED. N-TERMINAL  
CC LIES IN THE PERIPLASMIC SPACE.  
CC -1- MISCELLANEOUS: THE MUTANT DACA1191 IS STILL BINDING PENICILLIN,  
CC BUT IS BLOCKED IN THE RELEASE OF THE BOUND PENICILLOYL MOIETY; THE  
CC MUTANT ALSO FAILS TO CATALYSE THE D-ALANINE CARBOXYPEPTIDASE  
CC REACTION AS THE HYDROLYSIS OF THE ACYL-ENZYME FORMED WITH  
CC SUBSTRATE IS ALSO BLOCKED AND THE ACYL-ENZYME ACCUMULATES.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S11; ALSO KNOWN AS THE  
CC D-ALANYL-D-ALANINE CARBOXYPEPTIDASE I FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: X06479; CAA29774.1; -  
DR EMBL: M18276; AAA24553.1; -  
DR EMBL: AE000168; AAC73733.1; -  
DR EMBL: D90703; BAA35275.1; -  
DR EMBL: D90704; BAA35279.1; -  
DR EMBL: U82598; AAB40832.1; ALT\_INIT.  
DR EMBL: AE005242; AAC54966.1; -  
DR EMBL: AP002552; BAB34093.1; -  
DR EMBL: L07636; AAA66340.1; -  
DR PIR: A03420; ZPECP5.  
DR PIR: A28536; A28536.  
DR PIR: D28387; D28387.  
DR MEROPS: S11.003; -  
DR ECO2DBASE: H040.5; 6TH EDITION.  
DR Ecogene: EG10201; *dacA*.  
DR InterPro: IPR001967; *Ala/AlaCBPase1*.  
DR Pfam: PF00768; *Peptidase\_S11*; 1.  
DR PRINTS: PR00725; DADCBPASE1.  
KW Hydrolase; Carboxypeptidase; peptidoglycan synthesis; inner membrane;  
KW Cell wall; signal; complete proteome.  
FT SIGNAL 1 29  
FT CHAIN 30 403  
FT ACT\_SITE 73 73  
FT VARIANT 134 134 G->D (IN MUTANT DACA1191).  
FT CONFLICT 94 94 E->D (IN REF. 7).  
SO SEQUENCE 403 AA; 44444 MW; 7FAAB8E98452EF22 CRC64;  
  
Query Match 70.5%; Score 31; DB 1; Length 403;  
Best local similarity 75.0%; Pred. No. 59;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
QY 2 YLRALGK 9  
Db 165 YVNALGK 172  
  
RESULT 15  
GIDA\_CAMME  
ID GIDA\_CAMME STANDARD; PRT; 619 AA.  
AC Q9PNA7;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Glucose inhibited division protein A.  
GN GIDA OR CJ1188C.  
OS *Campylobacter jejuni*.  
OC Bacteria; Proteobacteria; epsilon subdivision; *Campylobacter* group;  
OC *Campylobacter*.  
OX NCBI\_TaxID=197;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NCTC 11168;  
RX MEDLINE-20150912; Pubmed-10688204;  
RA Parkhill J., Wren B.W., Mungall K., Kelsey J.M., Churcher C.,  
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,  
RA Jagers K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,  
RA Ouali M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,  
RA Whitehead S., Barrett B.G.,  
RT "The genome sequence of the food-borne pathogen *Campylobacter jejuni*  
RT reveals hypervariable sequences.",  
RL Nature 403:665-668(2000).  
[14]  
RP FUNCTION: NOT KNOWN.  
CC -1- SIMILARITY: BELONGS TO THE GIDA FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
 DR EMBL: AL139077; CAB73442.1; -  
 DR InterPro: IPR001327; FAD\_pyr\_redox.  
 DR InterPro: IPR002218; GIDA.  
 DR InterPro: IPR004416; GIDA\_sub.  
 DR InterPro: IPR002645; STAS.  
 DR Pfam: PF01134; GIDA.1.  
 DR PRINTS: PRO0368; FADPNR.  
 DR PRODOM: PD003738; GIDA.1.  
 DR TIGRFAMs: TIGR00136; gida.1.  
 DR PROSITE: PS01280; GIDA\_1; 1.  
 DR PROSITE: PS01281; GIDA\_2; FALSE\_NEG.  
 KW Complete proteome.  
 SQ SEQUENCE 619 AA; 69172 MW; 18261C2019852AE2 CRC64;

Query Match 70.5%; Score 31; DB 1; Length 619;  
 Best Local Similarity 75.0%; Pred. No. 90;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 Oy 2 YLRALGLK 9  
 11: 1111  
 Db 181 YLQTLGLK 188

Search completed: July 11, 2003, 11:55:16  
 Job time: 1.86071 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: July 11, 2003, 11:44:31 ; Search time 3.92931 Seconds  
(without alignments)  
471.946 Million cell updates/sec

Title: US-10-091-442-5  
Perfect score: 44  
Sequence: 1 OYLKALGLK 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues  
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_protist:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaeophages:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	79.5	270	16	P96690
2	35	79.5	802	11	O8VE90
3	35	79.5	808	4	O96A78
4	34	77.3	80	16	O9XOC1
5	34	77.3	187	10	O2A020
6	34	77.3	361	3	O9HDD6
7	34	77.3	363	2	O9AGZ2
8	34	77.3	406	16	O9RXM0
9	34	77.3	635	10	O9LDB5
10	34	77.3	768	12	O9IBW6
11	34	77.3	769	12	O9E6C8
12	34	77.3	1013	10	O9EXE9
13	34	75.0	228	16	O8XK60
14	33	75.0	301	16	O8R8F7
15	33	75.0	314	17	O8ZT63
16	33	75.0	348	5	O9GYX0

17	33	75.0	412	16	O9PE89	O9PE89 campylobact
18	33	75.0	413	16	P95236	P95236 mycobacteri
19	33	75.0	415	17	O96Z30	O96Z30 sulfolobus
20	33	75.0	1139	10	O48894	O48894 lactuca sat
21	33	75.0	1885	17	O8TOD1	O8TOD1 methanosarc
22	33	72.7	80	9	O03961	O03961 bacterioph
23	32	72.7	154	5	O9GRR9	O9GRR9 leishmania
24	32	72.7	176	11	O55155	O55155 rattus norv
25	32	72.7	194	5	O9V827	O9V827 drosophila
26	32	72.7	201	12	O9DS07	O9DS07 ascovirus d
27	32	72.7	239	17	O9HRD3	O9HRD3 halobacteri
28	32	72.7	274	3	O06688	O06688 saccharomyc
29	32	72.7	295	17	O8T509	O8T509 methanosarc
30	32	72.7	303	2	O93DS5	O93DS5 streptococc
31	32	72.7	309	16	O9KRM4	O9KRM4 bacillus ha
32	32	72.7	311	17	O976M4	O976M4 sulfolobus
33	32	72.7	327	2	O30552	O30552 helicobacte
34	32	72.7	327	16	O26096	O26096 helicobacte
35	32	72.7	327	16	O9ZJ34	O9ZJ34 helicobacte
36	32	72.7	342	5	O02093	O02093 caenorhabdi
37	32	72.7	342	16	O9HMD9	O9HMD9 pseudomonas
38	32	72.7	373	16	O8YET1	O8YET1 brucella me
39	32	72.7	408	5	O9UAT1	O9UAT1 caenorhabdi
40	32	72.7	422	16	O8Y266	O8Y266 raietonia s
41	32	72.7	447	16	O9WZ69	O9WZ69 thermotoga
42	32	72.7	481	4	O9NUP7	O9NUP7 homo sapien
43	32	72.7	506	16	O8REI1	O8REI1 fuscobacteri
44	32	72.7	509	2	O32444	O32444 vibrio chol
45	32	72.7	516	5	O22287	O22287 caenorhabdi

## ALIGNMENTS

RESULT 1  
P96690 PRELIMINARY: PRT: 270 AA.  
AC P96690:  
DT 01-MAY-1997 (TRMBLrel. 03, Created)  
DT 01-JAN-1998 (TRMBLrel. 05, Last sequence update)  
DT 01-MAR-2002 (TRMBLrel. 20, Last annotation update)  
DE YDFL protein.  
OS YDFL.  
OC Bacillus subtilis.  
CC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
OC Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RA MEDLINE=98044033; PubMed=9384377;  
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
RA Azevedo V., Bertoletto M.G., Bessieres P., Bolotin A., Borcherst S.,  
RA Broiliet S., Brunsch C.V., Caldwell B., Capuano V., Carter N.M.,  
RA Choi S.K., Codan J.J., Conneron I.F., Cummings N.J., Daniel R.A.,  
RA Dentzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
RA Entian K.D., Erlington J., Fabret C., Ferrari E., Foulger D.,  
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
RA Gilm S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
RA Guisepi G., Guy B.J., Hage K., Haiech J., Harwood C.R., Henaut A.,  
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
RA Joris B., Karamata D., Kasahara Y., Kletter-Blanchard M., Klein C.,  
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,

```

RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Nodack M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Paro V., Pohl T.M., Portelle D., Potwollit S., Prescott A.M.,
RA Presencan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivoita C., Roche E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
RA Sorokin A., Taconi E., Takagi T., Takahashi H., Takemaru K.,
RA Tsuchiuchi M., Tamakoshi A., Tanaka T., Terpsit P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wamburt R., Wedler E., Wedler H., Weitzenger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
DR EMBL: AB001488; BAA19380.1; -
DR EMBL: Z99106; CAB12353.1; -
DR InterPro: IPR000551; HTH_Merr.
DR Pfam: PF00376; merr; 1.
DR SMART: SM00422; HTH_MERR; 1.
DR DNA-binding: Transcription regulation: Complete proteome.
KW SSEQUENCE 270 AA; 31783 MW; 1C487674F59A6797 CRC64;
SQ
Query Match 79.5%; Score 35; DB 16; Length 270;
Best Local Similarity 87.5%; Pred. No. 38;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 1 QYRALGL 8
Db 56 KYRALGL 63

```

```

RA 08VE90
AC 08VE90;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical 88.4 kda protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC019504; AAH19504.1; -
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 12.
DR PRINTS: PR00320; GPROTEINBRPT.
DR SMART: SM00320; WD40; 13.
DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_2.
DR PROSITE: PS50082; WD_REPEATS_2; 9.
DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
KW Hypothetical protein
SQ SEQUENCE 802 AA; 88379 MW; EFC3CFB0425431B CRC64;
Query Match 79.5%; Score 35; DB 11; Length 802;
Best Local Similarity 87.5%; Pred. No. 1.2e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 1 QYRALGL 8
Db 1 QYRALGL 8

```

```

RA 096A78
AC 096A78;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical 89.0 kda protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-LUNG;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-CERVIX;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 13 WD REPEATS (TRP-ASP DOMAINS).
DR EMBL: BC014007; AAH14007.1; -
DR EMBL: BC010231; AAH10231.1; -
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 13.
DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_2.
DR PROSITE: PS50082; WD_REPEATS_2; 8.
DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
KW Hypothetical protein: Repeat; WD repeat.
SQ SEQUENCE 808 AA; 89034 MW; 01BDF217F8017030 CRC64;
Query Match 79.5%; Score 35; DB 4; Length 808;
Best Local Similarity 87.5%; Pred. No. 1.2e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 1 QYRALGL 8
Db 665 KYRALGL 672

```

```

RA 09X0C1
AC 09X0C1;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE GLUTAREDOXIN.
GN TM1031.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Uitterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima."
RL Nature 399:323-329(1999)
DR EMBL: AE001764; AAD36108.1; -
DR HSSP: P37687; IFOV.

```

DR TIGR; TM1031; -  
 DR InterPro: IPR002109; Glutaredoxin.  
 DR Pfam: PF00462; glutaredoxin; 1.  
 DR PRINTS: PR00160; GLUTAREDOXIN.  
 KM COMPLETE PROTEOME.  
 SQ SEQUENCE 80 AA; 9127 MW; B1802B9D3460262F CRC64;

Query Match 77.3%; Score 34; DB 16; Length 80;  
 Best Local Similarity 66.7%; Pred. No. 16;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 QYLRALGLK 9  
 :|:|:|:|  
 Db 22 EYFRSLGLK 30

## RESULT 5

ID 024020 PRELIMINARY; PRT; 187 AA.  
 AC 024020;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Cryptochrome (Fragment).  
 GN CYR2.  
 OS Lycopersicon esculentum (Tomato).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
 OX NCBI\_TaxID=4081;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kojimaoglu U., Schneider-Poesch H.A.;  
 RT "The evolution of cryptochrome genes."  
 RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ000695; CAA04247.1; -  
 DR InterPro: IPR005101; FAD\_binding\_7.  
 DR Pfam: PF03441; FAD\_binding\_7; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 187 187  
 SQ SEQUENCE 187 AA; 21093 MW; 26D5459DD510E737 CRC64;

Query Match 77.3%; Score 34; DB 10; Length 187;  
 Best Local Similarity 75.0%; Pred. No. 41;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 YLRALGLK 9  
 :|:|:|:|  
 Db 147 YLRALGLK 154

## RESULT 6

ID 09HDD6 PRELIMINARY; PRT; 361 AA.  
 AC 09HDD6;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE Mannitol-1-phosphate dehydrogenase.  
 GN MPD1.  
 OS Cryptococcus neoformans var. neoformans.  
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Heterobasidiomycetes;  
 OC Tremellomycetidae; Tremellales; Tremellaceae; Filobasidiella.  
 OX NCBI\_TaxID=40410;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-H99;  
 RC MEDLINE=20476444; PubMed=11021946;  
 RX Suvarna K., Bartiss A., Wong B.;  
 RA "Mannitol-1-phosphate dehydrogenase from Cryptococcus neoformans is a  
 zinc-containing long-chain alcohol/polyol dehydrogenase."  
 RT Microbiology 146:2705-2713(2000).  
 DR EMBL; AF186474; AAC17013.1; -

DR EMBL; AF175685; AAC09209.1; -  
 DR InterPro: IPR002085; adh\_zn\_family.  
 DR Pfam: PF00107; adh\_zinc; 1.  
 SQ SEQUENCE 361 AA; 37537 MW; F549926566CAF109 CRC64;

Query Match 77.3%; Score 34; DB 3; Length 361;  
 Best Local Similarity 66.7%; Pred. No. 84;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 QYLRALGLK 9  
 :|:|:|:|  
 Db 201 QYAKAMGLK 209

## RESULT 7

ID 09AGZ2 PRELIMINARY; PRT; 363 AA.  
 AC 09AGZ2;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Phosphoribosylaminoimidazole carboxylase.  
 GN PURK.  
 OS Ehrlichia canis.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Anaplasmataceae; Ehrlichia.  
 OX NCBI\_TaxID=944;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-OKLAHOMA;  
 RC MEDLINE=98371112; PubMed=9705412;  
 RX Ohashi N., Unver A., Zhl N., Rikihisa Y.;  
 RA "Cloning and characterization of multigenes encoding the  
 immunodominant 30-kilodalton major outer membrane proteins of  
 Ehrlichia canis and application of the recombinant protein for  
 serodiagnosis."  
 RT J. Clin. Microbiol. 36:2671-2680(1998).  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-OKLAHOMA;  
 RC MEDLINE=21153566; PubMed=11254561;  
 RX Ohashi N., Rikihisa Y., Unver A.;  
 RA "Analysis of Transcriptionally Active Gene Clusters of Major Outer  
 RT Membrane Protein Multigene Family in Ehrlichia canis and E.  
 RT chaffeensis."  
 RL Infect. Immun. 69:2083-2091(2001).  
 DR EMBL; AF324792; AAK31311.1; -  
 DR HSP; P09029; 1B6R.  
 DR InterPro: IPR003135; ATP-grasp.  
 DR Pfam: PF02222; ATP-grasp; 1.  
 DR TIGRfams; TIGR01161; purK; 1.  
 SQ SEQUENCE 363 AA; 40744 MW; 5296634FCF56DE75 CRC64;

Query Match 77.3%; Score 34; DB 2; Length 363;  
 Best Local Similarity 66.7%; Pred. No. 84;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 QYLRALGLK 9  
 :|:|:|:|  
 Db 115 QYRLMGLK 123

## RESULT 8

ID 09RXM0 PRELIMINARY; PRT; 406 AA.  
 AC 09RXM0;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Hypothetical protein DR0290.  
 GN DR0290.  
 OS Deinococcus radiodurans.  
 OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;

OC Deinococcaceae; Deinococcus.  
 OX NCBI\_TaxID=1299;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=R1;  
 RX MEDLINE=20036896; PubMed=10567266;  
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
 RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,  
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
 RA Fraser C.M.;  
 RT "Genome sequence of the radioresistant bacterium Deinococcus  
 RT radiodurans R1.";  
 RL Science 286:1571-1577(1999).  
 DR EMBL: AE001890; AAF09871.1; -.  
 DR TIGR: DR0290; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 406 AA; 42901 MW; 6B40D3F153F9FCB CRC64;

Query Match 77.3%; Score 34; DB 16; Length 406;  
 Best Local Similarity 100.0%; Pred. No. 95;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 YLRALGL 8  
 |||||  
 Db 33 YLRALGL 39

RESULT 9  
 Q9LDJ5 PRELIMINARY; PRT; 635 AA.  
 AC Q9LDJ5;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Cryochrome 2.  
 OS Lycopersicon esculentum (Tomato).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
 OX NCBI\_TaxID=4081;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20267445; PubMed=10809448;  
 RA Perrotta G., Ntutu L., Flamma F., Weller J.L., Kendrick R.E.,  
 RA Nebuloso E., Giuliano G.;  
 RT "Tomato contains homologues of Arabidopsis cryochromes 1 and 2.";  
 RL Plant Mol. Biol. 42:765-773(2000).  
 DR EMBL: AF130426; AAF72557.1; -.  
 DR EMBL: AF130425; AAF72556.1; -.  
 DR InterPro: IPR000474; DNA\_photolyase.  
 DR InterPro: IPR002081; DNA\_photolyase.  
 DR InterPro: IPR005101; DNA\_photolyase.  
 DR Pfam: PF00875; DNA\_photolyase; 1.  
 DR Pfam: PF03441; FAD\_binding\_7; 1.  
 DR PRINTS: PR00147; DNAPHOTLYASE.  
 DR PRODOM: PD004390; DNA\_photolyase; 1.  
 DR PROSITE: PS00394; DNA\_PHOTOLYASES\_1; 1.  
 SQ SEQUENCE 635 AA; 72249 MW; 937445E04DB86D6B CRC64;

Query Match 77.3%; Score 34; DB 10; Length 635;  
 Best Local Similarity 75.0%; Pred. No. 1.5e+02;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 YLRALGL 9  
 |||||  
 Db 284 YLRALGLR 291

RESULT 10  
 Q91BW6

ID Q91BW6 PRELIMINARY; PRT; 768 AA.  
 AC Q91BW6;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE UL8.  
 GN UL8.  
 OS Turkey herpesvirus.  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Marek's disease-like viruses.  
 OX NCBI\_TaxID=10390;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GA;  
 RX MEDLINE=92237304; PubMed=1315048;  
 RA Jones D., Lee L., Liu J.L., Kung H.J., Tiliotson J.K.;  
 RT "Marek disease virus encodes a basic-leucine zipper gene resembling  
 RT the fos/jun oncogenes that is highly expressed in lymphoblastoid  
 RT tumors.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:4042-4046(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GA;  
 RX MEDLINE=20283955; PubMed=10823954;  
 RA Lee L.F., Wu P., Sui D., Ren D., Kamil J., Kung H.J., Witter R.L.;  
 RT "The complete unique long sequence and the overall genomic  
 RT organization of the GA strain of Marek's disease virus.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:6091-6096(2000).  
 DR EMBL: AF147806; AAF66744.1; -.  
 DR InterPro: IPR002064; DNA\_pol\_B.  
 DR InterPro: IPR004996; Herpes\_HEPA.  
 DR Pfam: PF03324; Herpes\_HEPA; 1.  
 DR PROSITE: PS00116; DNA\_POLYMERASE\_B; UNKNOWN\_1.  
 SQ SEQUENCE 768 AA; 86859 MW; E4CA706C4D185C88 CRC64;

Query Match 77.3%; Score 34; DB 12; Length 768;  
 Best Local Similarity 66.7%; Pred. No. 1.9e+02;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 QYRALGL 9  
 |||||  
 Db 671 QYRALGL 679

RESULT 11  
 Q9E608 PRELIMINARY; PRT; 769 AA.  
 AC Q9E608;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE UL8 DNA helicase-primase associated protein-like protein.  
 GN MDV020.  
 OS Turkey herpesvirus.  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Marek's disease-like viruses.  
 OX NCBI\_TaxID=10390;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MD5;  
 RX MEDLINE=20392152; PubMed=10933706;  
 RA Tulman E.R., Afonso C.L., Lu Z., Zaak L., Rock D.L., Kutish G.F.;  
 RT "The genome of a very virulent Marek's disease virus.";  
 RL J. Virol. 74:7980-7988(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MD5;  
 RA Tulman E.R., Afonso C.L., Lu Z., Zaak L., Rock D.L., Kutish G.F.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF243438; AAG14200.1; -.  
 DR InterPro: IPR004996; Herpes\_HEPA.  
 DR Pfam: PF03324; Herpes\_HEPA; 1.  
 KW Helicase.

SO SEQUENCE 769 AA; 86907 MW; 28309BDBAF08333 CRC64;  
 Query Match 77.3%; Score 34; DB 12; Length 769;  
 Best Local Similarity 66.7%; Pred. No. 1.9e+02;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 OYLRALGK 9  
 11:11:11  
 Db 672 OYLRALGK 680

RESULT 12  
 Q9FKE9 PRELIMINARY; PRT; 1013 AA.  
 ID Q9FKE9  
 AC Q9FKE9  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Putative receptor-like serine/threonine kinase-partial protein  
 DE (Fragment).  
 DE F1N18.22.  
 GN Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Federici N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,  
 RA Alatafi H., Nguyen M., Lam B., Southwick A., Miranda M., Brooks S.,  
 RA Buehler E., Chao O., Chin C., Chlou J., Choi E., Gonzalez A.,  
 RA Hough B., Johnson-Hopson C., Khan S., Kim C., Koo T., Lee J.M.,  
 RA Lenz C., Liu A., Liu S., Mukherjee N., Pham P., Sakano H., Shin P.,  
 RA Tortum M., Vayenberg M., Yu G., Becker J., Theologis A., Davis R.W.,  
 RL Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AC008030; AAC10622.1; -  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR001611; LRR.  
 DR InterPro: IPR003592; LRR.  
 DR InterPro: IPR003591; LRR.  
 DR InterPro: IPR002290; Ser\_thr\_pkinase.  
 DR InterPro: IPR004040; STY\_pkinase.  
 DR InterPro: IPR001245; Tyr\_pkinase.  
 DR Pfam: PF00560; LRR; 7.  
 DR Pfam: PF00069; pkinase; 1.  
 DR PRINTS: PR00019; LEURICHRP.  
 DR ProDom: PD000001; Euk\_pkinase; 1.  
 DR SMART: SM00370; LRR; 5.  
 DR SMART: SM00369; LRR\_TYR; 2.  
 DR SMART: SM00221; STYK; 1.  
 DR SMART: SM00220; S\_TKC; 1.  
 DR SMART: SM00219; TYRK; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; 1.  
 KW Kinase; Receptor.  
 FT NON\_TER  
 SO SEQUENCE 1013 AA; 112631 MW; 6A326BE4B2B10215 CRC64;

Query Match 77.3%; Score 34; DB 10; Length 1013;  
 Best Local Similarity 77.8%; Pred. No. 2.6e+02;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 OYLRALGK 9  
 11:11:11  
 Db 647 OYLRALGK 655

RESULT 13  
 O8XK60 PRELIMINARY; PRT; 228 AA.  
 AC O8XK60  
 DT 01-MAR-2002 (TREMBLrel. 20, Created)

DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Dehydrobiocin synthetase.  
 GN BIOD OR CPE1543.  
 OS Clostridium perfringens.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;  
 OC Clostridiales; Clostridiaceae; Clostridium.  
 OC NCBI\_TaxID=1502;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=13 / TYPE A;  
 RX PubMed=11792842;  
 RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,  
 RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.,  
 RT "Complete genome sequence of Clostridium perfringens, an anaerobic  
 RT flesh-eater."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).  
 DR EMBL: AP003191; BAB81249.1; -  
 DR InterPro: IPR004472; BIOD.  
 DR TIGRFAMS: TIGR00347; BIOD; 1.  
 KW Complete proteome.  
 SO SEQUENCE 228 AA; 25668 MW; 336974B1360E727D CRC64;

Query Match 75.0%; Score 33; DB 16; Length 228;  
 Best Local Similarity 87.5%; Pred. No. 82;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 YLRALGK 9  
 11:11:11  
 Db 165 YLRALGK 172

RESULT 14  
 Q8R8F7 PRELIMINARY; PRT; 301 AA.  
 ID Q8R8F7  
 AC Q8R8F7  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Hypothetical protein TIE2045.  
 GN TIE2045.  
 OS Thermanaerobacter tengcongensis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;  
 OC Thermanaerobacteriales; Thermanaerobacteriaceae; Thermanaerobacter.  
 OC NCBI\_TaxID=119072;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MB4T / JCM11007;  
 RX MEDLINE=21992816; PubMed=11997336;  
 RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,  
 RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,  
 RA Tan H., Chen R., Wang J., Yu J., Yang H.,  
 RT "A complete sequence of T. tengcongensis genome."  
 RT Genome Res. 12:689-700(2002).  
 DR EMBL: AE013154; AAM25220.1; -  
 KW Hypothetical protein; Complete proteome.  
 SO SEQUENCE 301 AA; 35035 MW; 855F281CFCAD8297 CRC64;

Query Match 75.0%; Score 33; DB 16; Length 301;  
 Best Local Similarity 55.6%; Pred. No. 1.1e+02;  
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 OYLRALGK 9  
 11:11:11  
 Db 67 EYRSIGLK 75

RESULT 15  
 O8ZT63 PRELIMINARY; PRT; 314 AA.  
 AC O8ZT63  
 DT 01-MAR-2002 (TREMBLrel. 20, Created)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)

Dt	01-JUN-2002 (TREMBlrel. 21, Last annotation update)
De	Hypothetical protein PAE3418.
Gn	PAE3418.
Os	Pyrobaculum aerophilum.
Oc	Archaea: Crenarchaeota; Thermoprotei; Thermoproteales;
Ox	Thermoproteaceae; Pyrobaculum.
Ox	NCHI_TaxID=13773;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-IM2 / ATCC 51768 / DSM 7523;
RX	Pubmed=1192869;
RA	Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I., Miller J.H.;
RT	"Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum aerophilum".
RL	Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
DR	EMBL; AE009929; AAL64900.1; -.
DR	IamPro; IPR001279; Blactnase_1like.
KW	Interf; PF00753; lactnase_B; 1. Hypothetical protein; Complete proteome.
SO	SEQUENCE 314 AA; 34125 MW; 9B22EBB710FD455 CRC64;
Oy	Query Match            75.0%; Score 33; DB 17; Length 314; Best Local Similarity   66.7%; Pred. NO. 1.2e+02; Matches     6; Conservative       2; Mismatches     1; Indels     0; Gaps     0;
Db	1 OYLALGLX 9 :   :    292 RYLRKIGLK 300

Search completed: July 11, 2003, 11:59:00  
Job time : 6.92931 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:42:16 ; Search time 11.4761 Seconds  
(without alignments)  
232.223 Million cell updates/sec

Title: US-10-091-442-6  
Perfect score: 98  
Sequence: 1 ADLSGIASGRLYISRMWCK 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: A\_Geneseq\_101002.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	92	93.9	20	15	AA857112 Human megakaryocyt
2	90	91.8	380	15	AA848379 Human megakaryocyt
3	90	91.8	380	20	AA108254 Human megaslin prote
4	90	91.8	380	21	AA824142 Human megaslin prote
5	90	91.8	380	22	AA83075 Human megaslin prote
6	83	84.7	368	20	AA108256 Mouse megaslin prote
7	83	84.7	368	21	AA824151 Mouse megaslin prote
8	83	84.7	368	22	AA83077 Murine megaslin prote
9	83	84.7	380	20	AA108255 Rat megaslin prote
10	83	84.7	380	21	AA824150 Rat megaslin prote

11	83	84.7	380	22	AA64286
12	83	84.7	380	22	AA83076
13	50	51.0	615	22	AAU28187
14	50	51.0	615	22	AA83480
15	47	48.0	394	22	AA853280
16	47	48.0	395	22	AA847207
17	47	48.0	405	22	AA101601
18	47	48.0	617	22	ABG28280
19	46	46.9	327	22	AA89057
20	46	46.9	327	22	AA831515
21	46	46.9	618	22	ABG24021
22	45	45.9	307	22	AA692577
23	45	45.9	373	23	AB87879
24	45	45.9	387	23	AAU78358
25	45	45.9	406	23	AAU78357
26	44.5	45.4	434	12	AA813483
27	44	44.9	377	22	AA848382
28	44	44.9	453	23	AB848359
29	44	44.9	453	23	AA72661
30	43	43.9	135	22	AA72661
31	43	43.9	316	20	AA72661
32	43	43.9	391	20	AA741029
33	43	43.9	391	20	AA725926
34	43	43.9	391	21	AA811314
35	43	43.9	391	21	AA810953
36	43	43.9	391	21	AA810954
37	43	43.9	391	22	AA72651
38	43	43.9	391	23	AB874946
39	43	43.9	392	19	AAW48391
40	43	43.9	392	20	AAV04120
41	43	43.9	392	23	AAE14266
42	43	43.9	400	20	AAV41030
43	43	43.9	400	21	AA811315
44	43	43.9	400	22	AA72652
45	43	43.9	400	23	AB874947

#### ALIGNMENTS

RESULT 1	AA857112	standard; peptide: 20 AA.
XX	AA857112:	
AC	AA857112:	
XX	16-AUG-1994 (first entry)	
DT		
XX		
DE	Human megakaryocyte differentiation factor peptide 6.	
XX		
KW	Human megakaryocyte differentiation factor; MGF; thrombopoietin;	
KW	hematopoietic stimulating factor; thrombocytopoietin; platelet;	
KW	bone marrow transplantation; cancer chemotherapy.	
XX		
OS	Homo sapiens.	
XX		
FT	Key	Location/Qualifiers
FT	Misc-difference 18	/note= "not determined"
PN	EP583884-A.	
XX		
PD	23-FEB-1994.	
XX		
XX	19-JUL-1993;	93EP-0305654.
FE		
XX	17-JUL-1992;	92JP-0212305.
PR	04-MAR-1993;	93JP-0067339.
XX		
PA	(SUNR) SUNTORY LTD.	
PA	(TSUJ) TSUJIMOTO M.	
XX		
PI	Ishida N, Iwasa F, Kurihara T, Miura K, Nakazato H;	

Rat megaslin protein  
Rat megaslin protein  
Novel human secret  
Human protein sequ  
Human polypeptide  
Human NOVI protein  
Protein encoded by  
Novel human diago  
Canine GnRH recept  
A dog gonadotropin  
Novel human diago  
C glutamicum prote  
Amino acid sequenc  
Modified human pro  
Human protein C in  
P. denitrificans CO  
I. ricinus salivari  
Modified human pro  
Listeria monocytog  
Human headpin (for  
Human serine prote  
Human lung tumor a  
Human huplin prote  
Human lung cancer-  
Human serine prote  
Human serine prote  
Human headpin (for  
Human lung tumour  
Human lung tumour  
Pancrreas derived p  
Human Pancreas-der  
Human lung tumor a  
Human lung cancer-  
Human headpin (hea  
Human lung tumour

PI Tsujimoto M, Tsuruoka N, Yamaguchi N, Yamachi K;  
 XX  
 DR WPI: 1994-058782/08.  
 XX  
 PT New megakaryocyte differentiation factor - isolated from human  
 PT epidermoid carcinoma cells, used to treat conditions involving a  
 PT decrease in platelets  
 XX  
 PS Claim 1; Page 20; 47pp: English.  
 XX  
 CC Human MDF (see AAR48379) can be isolated from a culture of human  
 CC epidermoid carcinoma A431 cells in protein-free medium. The MDF  
 CC stimulates differentiation of megakaryocytes from myeloid cells  
 CC in the presence of IL-3. The MDF acts in vivo as a thrombopoietin  
 CC making it useful for treatment of diseases involving a decrease  
 CC in platelet number (esp. thrombocytopaenia) such as occurs in bone  
 CC marrow transplantation and in chemotherapy. MDF has mol.wt.  
 CC 55-57kD by SDS-PAGE, exhibits an isoelectric point of 6.0-7.0 and  
 CC contains an amino acid sequence comprising at least one of the  
 CC sequences AAR57107-R57115.  
 XX  
 SQ Sequence 20 AA;  
 Query Match 93.9%; Score 92; DB 15; Length 20;  
 Best Local Similarity 95.0%; Pred. No. 5e-08;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 ADLSGIASGGRLYISRMKG 20  
 DB 1 ADLSGIASGGRLYISRMKG 20  
 RESULT 2  
 AAR48379  
 ID AAR48379 standard; Protein: 380 AA.  
 XX  
 AC AAR48379;  
 XX  
 DT 16-AUG-1994 (first entry)  
 XX  
 DE Human megakaryocyte differentiation factor.  
 XX  
 KM Human megakaryocyte differentiation factor; MDF; thrombopoietin;  
 KM hematopoietic stimulating factor; thrombocytopaenia; platelet;  
 KM bone marrow transplantation; cancer chemotherapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP583884-A.  
 XX  
 PD 23-FEB-1994.  
 XX  
 PF 19-JUL-1993; 93EP-0305654.  
 XX  
 PR 17-JUL-1992; 92JP-0212305.  
 PR 04-MAR-1993; 93JP-0067339.  
 XX  
 PA (SUNR) SUNTORY LTD.  
 PA (TSUJ) TSUJIMOTO M.  
 XX  
 PI Iahida N, Iwasa F, Kurihara T, Miura K, Nakazato H;  
 PI Tsujimoto M, Tsuruoka N, Yamaguchi N, Yamachi K;  
 XX  
 DR WPI: 1994-058782/08.  
 DR N-PSDB; AAO56670.  
 XX  
 PT New megakaryocyte differentiation factor - isolated from human  
 PT epidermoid carcinoma cells, used to treat conditions involving a  
 PT decrease in platelets  
 XX  
 PS Claim 7; Page 30-32; 47pp: English.  
 XX  
 CC Human MDF can be isolated from a culture of human epidermoid

CC carcinoma A431 cells in protein-free medium. The MDF stimulates  
 CC differentiation of megakaryocytes from myeloid cells in the presence  
 CC of IL-3. The MDF acts in vivo as a thrombopoietin making it useful  
 CC for treatment of diseases involving a decrease in platelet number  
 CC (esp. thrombocytopaenia) such as occurs in bone marrow  
 CC transplantation and in chemotherapy. MDF has mol.wt. 55-57kD by  
 CC SDS-PAGE and exhibits an isoelectric point of 6.0-7.0.  
 XX  
 SQ Sequence 380 AA;  
 Query Match 91.8%; Score 90; DB 15; Length 380;  
 Best Local Similarity 95.0%; Pred. No. 2.7e-06;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 ADLSGIASGGRLYISRMKG 20  
 DB 305 ADLSGIASGGRLYISRMKG 324  
 RESULT 3  
 AAY08254  
 ID AAY08254 standard; Protein: 380 AA.  
 XX  
 AC AAY08254;  
 XX  
 DT 14-JUL-1999 (first entry)  
 XX  
 DE Human megasin protein.  
 XX  
 KM Megasin; mesangial cell; treatment; diagnosis; disease; Iga nephropathy;  
 KM human; rat; murine.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO915652-A1.  
 XX  
 PD 01-APR-1999.  
 XX  
 PF 22-SEP-1998; 98WO-JP04269.  
 XX  
 PR 22-SEP-1997; 97JP-0275302.  
 XX  
 PA (KURO) KUROKAWA K.  
 PA (MIYA) MIYATA T.  
 XX  
 PI Miyata T;  
 XX  
 DR WPI: 1999-276983/23.  
 DR N-PSDB; AAX56712.  
 XX  
 PT Megasin protein expressed specifically in mesangial cells  
 PT Claim 1; Page 62-64; 100pp; Japanese.  
 XX  
 PS This invention describes the isolation of novel megasin nucleic acid and  
 PS proteins from human, rat and mouse tissue. This protein is expressed  
 CC specifically in mesangial cells. The products of the invention are  
 CC useful for the treatment and diagnosis of diseases involving mesangial  
 CC cells, such as Iga nephropathy.  
 XX  
 SQ Sequence 380 AA;  
 Query Match 91.8%; Score 90; DB 20; Length 380;  
 Best Local Similarity 95.0%; Pred. No. 2.7e-06;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 ADLSGIASGGRLYISRMKG 20  
 DB 305 ADLSGIASGGRLYISRMKG 324  
 RESULT 4  
 AAB24142



ID AAB24142 standard; Protein: 380 AA.  
 XX  
 AC AAB24142;  
 XX  
 DT 30-JAN-2001 (first entry)  
 XX  
 DE Human megalin protein sequence SEQ ID NO:2.  
 XX  
 KW Megalin: mesangium-predominant gene; serpin regulated; nephropathy;  
 KW Iga; immunoglobulin A; detection; renal function; renal disorder;  
 KW diagnosis; biological sample; blood; urine.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W0200057189-A1.  
 XX  
 PD 28-SEP-2000.  
 XX  
 PF 17-MAR-2000; 2000MO-JP01646.  
 XX  
 PR 19-MAR-1999; 99JP-0075305.  
 PR 28-OCT-1999; 99JP-0306623.  
 XX  
 PA (KURO/) KUROKAWA K.  
 PA (FUSO ) FUSO PHARM IND LTD.  
 PA (MIYA/) MIYATA T.  
 XX  
 PI Miyata T;  
 XX  
 DR WPI: 2000-611642/58.  
 DR N-PSDB: AAA99294.  
 XX  
 PT Evaluating renal function comprises assaying megalin protein in  
 PT biological sample -  
 XX  
 PS Example 2; Page 66-69; 93pp; Japanese.  
 XX  
 CC The present invention describes a method for evaluating renal function.  
 CC The method comprises assaying megalin protein in biological sample. Also  
 CC described are: (1) use of an anti-megalin protein antibody for diagnosing  
 CC renal function; and (2) a kit for detecting megalin protein comprising:  
 CC (a) anti-megalin protein antibody attached to solid magnetic particles;  
 CC (b) direct or indirect fixing for the antibody to the particles; and  
 CC (c) a magnet. The process is useful for evaluating renal function and  
 CC diagnosing renal disorders by assaying megalin protein in biological  
 CC samples (preferably urine or blood). The process is reproducible and  
 CC gives accurate results. The present sequence represents the human megalin  
 CC protein, which is given in the exemplification of the present invention.  
 XX  
 SQ Sequence 380 AA;  
 XX  
 QY Query Match 91.8%; Score 90; DB 21; Length 380;  
 Best Local Similarity 95.0%; Pred. No. 2,7e-06;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 XX  
 DB 1 ADLSGIASGRLYISRMGMK 20  
 ||||||||||||||||  
 305 ADLSGIASGRLYISRMGMK 324

OS Homo sapiens.  
 XX  
 PN W0200124628-A1.  
 XX  
 PD 12-APR-2001.  
 XX  
 PF 06-OCT-2000; 2000MO-JP06988.  
 XX  
 PR 06-OCT-1999; 99JP-0285736.  
 XX  
 PA (KURO/) KUROKAWA K.  
 PA (MIYA/) MIYATA T.  
 XX  
 PI Miyata T;  
 XX  
 DR WPI: 2001-300136/31.  
 DR N-PSDB: AAF82438.  
 XX  
 PT Mouse model for mesangial cell proliferative nephritis for development  
 PT and screening of new treatments -  
 XX  
 PS Example 4; Page 44-46; 62pp; Japanese.  
 XX  
 CC The present sequence is human megalin. The human megalin coding  
 CC sequence may be introduced into a mouse to produce an animal model of  
 CC mesangial cell proliferative nephritis. The symptoms include  
 CC enlargement of the mesangial base region, sedimentation of an immune  
 CC complex and an increase in mesangial cells. The animal model is useful  
 CC for analysing the pathology of chronic glomerular diseases and for  
 CC screening compositions for prevention and treatment of the diseases.  
 CC Highly uniform models can be made easily and in large numbers using  
 CC this method.  
 XX  
 SQ Sequence 380 AA;  
 XX  
 QY Query Match 91.8%; Score 90; DB 22; Length 380;  
 Best Local Similarity 95.0%; Pred. No. 2,7e-06;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 XX  
 DB 1 ADLSGIASGRLYISRMGMK 20  
 ||||||||||||||||  
 305 ADLSGIASGRLYISRMGMK 324

RESULT 6  
 ID AAY08256 standard; Protein: 368 AA.  
 XX  
 AC AAY08256;  
 XX  
 DT 14-JUL-1999 (first entry)  
 XX  
 DE Mouse megalin protein.  
 XX  
 KW Megalin: mesangial cell; treatment; diagnosis; disease; Iga nephropathy;  
 KW human; rat; murine.  
 XX  
 OS Mus musculus.  
 XX  
 PN W09915652-A1.  
 XX  
 PD 01-APR-1999.  
 XX  
 PF 22-SEP-1998; 98MO-JP04269.  
 XX  
 PR 22-SEP-1997; 97JP-0275302.  
 XX  
 PA (KURO/) KUROKAWA K.  
 PA (MIYA/) MIYATA T.

Key Location/Qualifiers  
 FT Protein 1..368  
 FT /note="partial sequence"

XX Miyata T;  
PI  
XX WPI; 1999-276983/23.  
DR N-PSDB; AAX56714.  
XX  
XX Megsin protein expressed specifically in mesangial cells  
PT  
XX Claim 1; Page 76-79; 100pp; Japanese.  
PS  
XX This invention describes the isolation of novel megin nucleic acid and  
CC proteins from human, rat and mouse tissue. This protein is expressed  
CC specifically in mesangial cells. The products of the invention are  
CC useful for the treatment and diagnosis of diseases involving mesangial  
CC cells, such as IgA nephropathy.  
SQ Sequence 368 AA;

Query Match 84.7%; Score 83; DB 20; Length 368;  
Best Local Similarity 80.0%; Pred. No. 3.4e-05;  
Matches 16; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
QY 1 ADLSGIASGRLYSRMGX 20  
| | | | | | | | | | : : : |  
Db 293 ADLSGIASGRLYSKLMX 312

RESULT 7  
AAB24151  
ID AAB24151 standard; Protein; 368 AA.  
XX  
AC AAB24151;  
XX  
DT 30-JAN-2001 (first entry)  
XX  
DE Mouse megin protein sequence SEQ ID NO:21.  
XX  
KW Megsin; mesangium-predominant gene; serpin regulated; nephropathy;  
KW IgA; immunoglobulin A; detection; renal function; renal disorder;  
KW diagnosis; biological sample; blood; urine.  
XX  
MS Mus musculus.

WO200057189-A1.  
PN  
PD 28-SEP-2000.  
PF 17-MAR-2000; 2000WO-JP01646.  
PR 19-MAR-1999; 99JP-0075305.  
PR 28-OCT-1999; 99JP-0306623.  
XX  
PA (KURO/) KURUKAWA K.  
PA (FUSO) FUSO PHARM IND LTD.  
PA (MIYA/) MIYATA T.  
PI  
PI Miyata T;  
DR WPI; 2000-611642/58.  
DR N-PSDB; AAC55239.

Evaluating renal function comprises assaying megin protein in  
biological sample

Disclosure; Page 89-91; 93pp; Japanese.

The present invention describes a method for evaluating renal function.  
The method comprises assaying megin protein in biological sample. Also  
described are: (1) use of an anti-megsin protein antibody for diagnosing  
renal function; and (2) a kit for detecting megin protein comprising:  
(a) anti-megsin protein antibody attached to solid magnetic particles;  
(b) direct or indirect fixing for the antibody to the particles; and  
(c) a magnet. The process is useful for evaluating renal function and

Query Match	Best Local Similarity	Score	DB	Length
Matches 16: Conservative	84.7%;	80.0%;	3;	Indels 0; Gaps 0;
1 ADLSCIASGGRLYISRMGMK 20				
293 ADLSCIASGGRLYISRMGMK 312				
Sequence 368 AA;				
Query Match	Best Local Similarity	Score	DB	Length
Matches 16: Conservative	84.7%;	80.0%;	3;	Indels 0; Gaps 0;
1 ADLSCIASGGRLYISRMGMK 20				
293 ADLSCIASGGRLYISRMGMK 312				
Sequence 368 AA;				

```

RESULT 9
AA08255
ID AA08255 standard; Protein: 380 AA.
XX
XX
AC AA08255;
XX
DT 14-JUL-1999 (first entry)
XX
DE Rat megalin protein.
XX
KW Megalin; mesangial cell; treatment; diagnosis; disease; Iga nephropathy;
XX human; rat; murine.
XX
OS Rattus rattus.
XX
PM WO9915652-A1.
XX
PD 01-APR-1999.
XX
PE 22-SEP-1998; 98WO-JP04269.
XX
PR 22-SEP-1997; 97JP-0275302.
XX
PA (KUROO/) KUROKAWA K.
XX (MIYA/) MIYATA T.
XX
PI Miyata T;
XX
DR WPI: 1999-276983/23.
DR N-PSDB: AAX56712.
XX
XX
PT Megalin protein expressed specifically in mesangial cells
XX
XX
PS Claim 1; Page 69-72; 100pp: Japanese.
XX
CC This invention describes the isolation of novel megalin nucleic acid and
CC proteins from human, rat and mouse tissue. This protein is expressed
CC specifically in mesangial cells. The products of the invention are
CC useful for the treatment and diagnosis of diseases involving mesangial
CC cells, such as Iga nephropathy.
XX
SQ Sequence 380 AA;

Query Match 84.7%; Score 83; DB 20; Length 380;
Best Local Similarity 80.0%; Pred. No. 3.6e-05;
Matches 16; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 ADLSGIASGGRLYSRMMGK 20
   |||||||:|:|:|
DB 305 ADLSGIASGGRLYSKLMHK 324

RESULT 10
AAB24150
ID AAB24150 standard; Protein: 380 AA.
XX
XX
AC AAB24150;
XX
DT 30-JAN-2001 (first entry)
XX
DE Rat megalin protein sequence SEQ ID NO:19.
XX
KW Megalin; mesangium-predominant gene; serpin regulated; nephropathy;
XX Iga; immunoglobulin A; detection; renal function; renal disorder;
XX diagnosis; biological sample; blood; urine.
XX
OS Rattus norvegicus.
XX
PM Key Location/Qualifiers
FH Misc-difference 51
FT /note= "unspecified"
FT Misc-difference 94
FT /note= "unspecified"

```

```

XX      WO200057189-A1.
PN      XX
PD      28-SEP-2000.
PE      17-MAR-2000; 2000WO-JP01646.
XX      XX
XX      19-MAR-1999; 99JP-0075305.
PR      28-OCT-1999; 99JP-0306623.
XX      XX
PA      (KUROO/) KUROKAWA K.
PA      (FUSO ) FUSO PHARM IND LTD.
PA      (MIYA/) MIYATA T.
PI      Miyata T;
DR      WPI: 2000-611642/58.
DR      N-PSDB: AAC55238.
XX      XX
PT      Evaluating renal function comprises assaying mesgin protein in
PT      biological sample -.
PS      Example 2: Page 81-84; 93pp; Japanese.
XX      XX
XX      The present invention describes a method for evaluating renal function.
CC      The method comprises assaying mesgin protein in biological sample. Also
CC      described are: (1) use of a anti-mesgin protein antibody for diagnosing
CC      renal function; and (2) a kit for detecting mesgin protein comprising:
CC      (a) anti-mesgin protein antibody attached to solid magnetic particles;
CC      (b) direct or indirect fixing for the antibody to the particles; and
CC      (c) a magnet. The process is useful for evaluating renal function and
CC      diagnosing renal disorders by assaying mesgin protein in biological
CC      samples (preferably urine or blood). The process is reproducible and
CC      gives accurate results. The present sequence represents the rat mesgin
CC      protein, which is given in the exemplification of the present invention.
XX      XX
SO      Sequence      380 AA;
XX      XX
XX      Query Match      84.7%; Score 83; DB 21; Length 380;
XX      Best Local Similarity 80.0%; Pred. No. 3.6e-05;
XX      Matches 16; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
OY      1 ADLSGIASGGRLYISRMNGK 20
        |||||||||||:|:|
DB      305 ADLSGIASGGRLYVSKLMHK 324
XX      XX
RESULT 11
AA64286
ID      AA64286 standard; Protein; 380 AA.
XX      XX
AC      AA64286;
XX      XX
DT      21-SEP-2001 (first entry)
XX      XX
DE      Rat mesgin protein.
XX      XX
KW      Rat; mesgin; renal mesangial cell; mesangium proliferative nephritis.
XX      XX
OS      Rattus norvegicus.
XX      XX
PN      WO200148019-A1.
XX      XX
PD      05-JUL-2001.
XX      XX
PF      26-DEC-2000; 2000WO-JP09251.
XX      XX
PR      28-DEC-1999; 99JP-0373677.
XX      XX
PA      (KUROO/) KUROKAWA K.
PA      (MIYA/) MIYATA T.
PI      Miyata T;

```

```

XX DR WPI: 2001-425651/45.
XX DR N-PSDB: AAH48181.
XX PT New antibody recognizing a partial sequence of rat mesgin protein for
XX PT diagnosis of mesangium proliferative nephritis
XX PS Disclosure: Page 54-56; 63pp; Japanese.
XX CC The present invention relates to a novel antibody which recognises a
XX CC peptide consisting of residues 341-354 of rat mesgin protein. The present
XX CC sequence is the protein sequence for rat mesgin, which was used in the
XX CC present invention. Mesgin is highly expressed in renal mesangial cells
XX CC and its level is elevated in mesangium proliferative nephritis. Assay of
XX CC the serum or urine level using the antibody is therefore indicative of
XX CC this type of disorder.
XX SQ Sequence 380 AA:

Query Match      84.7%; Score 83; DB 22; Length 380;
Best Local Similarity 80.0%; Pred. No. 3.6e-05;
Matches 16; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 ADLSGIASGRLYISRMWGK 20
   |||||
Db 305 ADLSGIASGRLYISKLMHK 324

RESULT 12.
ID AAB83076 standard; Protein; 380 AA.
AC AAB83076;
XX 10-JUL-2001 (first entry)
DT
XX Rat mesgin protein.
DE
XX Rat; mesgin: mesangial cell proliferative nephritis; nephrotropic;
KW transgenic mouse; glomerular disease; animal model; drug screening.
XX
XX Rattus norvegicus.
XX WO200124628-A1.
XX PN
XX 12-APR-2001.
XX PD
XX 06-OCT-2000; 2000WO-JP06988.
XX PF
XX 06-OCT-1999; 99JP-0285736.
XX PR
XX (KUROO) KUROKAWA K.
XX PA (MIYA) MIYATA T.
XX PI Miyata T;
XX WPI: 2001-300136/31.
XX DR N-PSDB: AAF82439.
XX PT Mouse model for mesangial cell proliferative nephritis for development
XX PT and screening of new treatments
XX PS Disclosure: Page 48-50; 62pp; Japanese.
XX CC The present sequence is rat mesgin. The human mesgin coding
XX CC sequence may be introduced into a mouse to produce an animal model of
XX CC mesangial cell proliferative nephritis. The symptoms include
XX CC enlargement of the mesangial base region, sedimentation of an immune
XX CC complex and an increase in mesangial cells. The animal model is useful
XX CC for analysing the pathology of chronic glomerular diseases and for
XX CC screening compositions for prevention and treatment of the diseases.
XX CC Highly uniform models can be made easily and in large numbers using
XX CC this method.

```

```

XX SQ Sequence 380 AA:

Query Match      84.7%; Score 83; DB 22; Length 380;
Best Local Similarity 80.0%; Pred. No. 3.6e-05;
Matches 16; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 ADLSGIASGRLYISRMWGK 20
   |||||
Db 305 ADLSGIASGRLYISKLMHK 324

RESULT 13
ID AAU28187 standard; Protein; 615 AA.
AC AAU28187;
XX 18-DEC-2001 (first entry)
DT
XX Novel human secretory protein, Seq ID No 356.
XX DE
XX
XX Human; secreted protein; arthritis; Crohn's disease; sepsis; shock;
XX KW ischaemia-reperfusion injury; haematopoiesis; cancer; neuropathy;
XX KW transgenic animal; Alzheimer's disease; Parkinson's disease; burn;
XX KW amyotrophic lateral sclerosis; platelet disorder; thrombocytopenia;
XX KW ulcer; osteoporosis; bone degenerative disorder; periodontal disease;
XX KW gut protection; lung; liver fibrosis; immune deficiency; infection;
XX KW severe combined immunodeficiency; SCID; autoimmune disorder; allergy;
XX KW multiple sclerosis; rheumatoid arthritis; diabetes mellitus; asthma;
XX KW fertility; analgesic; pain; antigen.
XX
XX Homo sapiens.
XX OS
XX WO200166689-A2.
XX PN
XX 13-SEP-2001.
XX PD
XX 05-MAR-2001; 2001WO-US04942.
XX PF
XX 07-MAR-2000; 2000US-0519705.
XX PR 19-MAY-2000; 2000US-0574454.
XX PR 17-JUN-2000; 2000US-0596193.
XX PR 14-JUL-2000; 2000US-0616847.
XX PR 19-SEP-2000; 2000US-0665363.
XX PR 20-OCT-2000; 2000US-0693267.
XX
XX (HXSE-) HXSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Xu C, Wehrman T, Ren F, Ma Y, Zhou P;
XX PI Zhao QA, Yang Y, Drmanac RT, Zhang J, Chen R, Xue AJ, Wang J;
XX DR N-PSDB: AAS45087.
XX DR WPI: 2001-589934/66.
XX PT Novel polypeptides and nucleic acids obtained from cDNA libraries
XX PT prepared from various human tissues, for diagnosis and treatment of
XX PT cancer, neurological, inflammatory, and autoimmune disorders
XX PS Example 5; SEQ ID No 356; 107pp; English.
XX CC The invention relates to novel isolated human secreted polypeptides (I)
XX CC and polynucleotides (II). (I) and (II) are useful for treating
XX CC inflammatory conditions such as arthritis, nephritis, Crohn's disease,
XX CC ischaemia-reperfusion injury, shock, sepsis, immune responses, and is
XX CC involved in increasing haematopoiesis, stem cell survival, bone growth
XX CC and remodeling. (I), (II) and modulators of (II) are useful for
XX CC prophylaxis or treatment of one or more cancers. (II) is also useful for
XX CC creating transgenic animals useful for studying the in vivo activities of
XX CC the polypeptide as well as for studying modulators of the polypeptides.
XX CC (I) induces the proliferation of neural cells and regeneration of nerve
XX CC and brain tissue and is useful for the treatment of central and
XX CC peripheral nervous system diseases and neuropathies, such as Alzheimer's,

```

CC Parkinson's disease, Huntington's disease, and amyotrophic lateral  
CC sclerosis. In addition, (1) is involved in chemotactic or chemokinetic  
CC activity, regulation of hematopoiesis and is useful for treating myeloid  
CC or lymphoid cell disorders, platelet disorders such as thrombocytopenia  
CC and for regeneration of bone, cartilage, tendon, ligament and/or nerve  
CC tissue growth, and in tissue repair, healing of burns, incisions,  
CC ulcers, for treating osteoporosis, osteoarthritis, bone degenerative  
CC disorders, or periodontal disease. Furthermore, (1) is also useful for  
CC gut protection or regeneration and treatment of lung or liver fibrosis,  
CC reperfusion injury in various tissues, various immune deficiencies and  
CC disorders including severe combined immunodeficiency (SCID), bacterial or  
CC fungal infections, autoimmune disorders e.g. multiple sclerosis,  
CC rheumatoid arthritis, diabetes mellitus, myasthenia gravis, allergic  
CC reactions and conditions, such as asthma or other respiratory problems.  
CC In addition, (1) affects biorhythms or circadian cycles of rhythms,  
CC fertility, metabolism, catabolism, anabolism, storage or elimination of  
CC dietary fat, lipid, protein, carbohydrate, vitamins, minerals, provides  
CC analgesic effects or other pain reducing effects, immunoglobulin like  
CC activity and can act as an antigen in a vaccine composition to raise an  
CC immune response. AAU28020-AAU28395 represent novel human secreted protein  
CC amino acid sequences of the invention.

XX Sequence 615 AA;

Query Match. 51.0%; Score 50; DB 22; Length 615;  
Best Local Similarity 83.3%; Pred. No. 13;  
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 4 SGASGRLYIS 15

DB 454 AGAASGRLYIS 465

## RESULT 14

AAB93480

ID AAB93480 standard; Protein: 615 AA.

XX AAB93480;

XX 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:12768.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI: 2001-318749/34.

XX Claim 8; SEQ ID 12768; 2537pp + CD ROM; English.

CC The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.

XX Sequence 615 AA;

Query Match. 51.0%; Score 50; DB 22; Length 615;  
Best Local Similarity 83.3%; Pred. No. 13;  
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 4 SGASGRLYIS 15

DB 454 AGAASGRLYIS 465

## RESULT 15

AAB53280

ID AAB53280 standard; Protein: 394 AA.

XX AAB53280;

XX 12-FEB-2002 (first entry)

XX Human polypeptide #20.

XX Human; nootropic; neuroprotective; anticonvulsant; antidepressant;

XX neuroleptic; tranquiliser; antiarrhythmic; cardiac; antidiabetic;

XX antiinflammatory; antihypertensive; hepatotropic; virucide; antidiabetic;

XX nephrotropic; anorectic; cytostatic; vaccine; neurological disease;

XX cardiovascular disease; respiratory disease; liver disease;

XX renal disease; skeletal muscle disease; gastrointestinal disease;

XX placental disease; testicular cancer; male fertility; pancreatic disease.

XX Homo sapiens.

XX WO200181363-A1.

XX 01-NOV-2001.

XX 26-APR-2001; 2001WO-US13360.

XX 27-APR-2000; 2000US-199963P.

XX 11-MAY-2000; 2000US-203336P.

XX 25-MAY-2000; 2000US-207087P.

XX 26-MAY-2000; 2000US-207546P.

XX (SMIK ) SMITHKLINE BEECHAM CORP.

XX (SMIK ) SMITHKLINE BEECHAM PLC.

XX Agrawal P, Murdock PR, Rizvi SK, Smith RF, Xiang Z, Kabnick KS;

XX Lai Y, Xie Q.

XX WPI: 2002-041392/05.

XX N-PSDB: ABA90345.

XX Novel polypeptides and polynucleotides useful as a vaccine for preventing and treating diseases associated with the polypeptide, e.g. Alzheimer's disease, dyslipidemia, obesity, diabetes, infertility PT asthma, amesias -

PS Claim 1: Page 90-91; 116pp; English.

xx The invention relates to an isolated polypeptide comprising a 277, 480, 583, 581, 628, 424, 638, 229, 310, 841, 241, 369, 382, 185, 586, 1026, 844, 782, 262, 394, 471, 485, 286, 533, 495, 330, 619, 450, 462, 255, 784, 252, 593, 472, 607, 781, 640, 686 or 154 amino acid sequence as given in the specification. The polypeptides, modulators of the polypeptides and antibodies against the polypeptides are useful for treating diseases such as neurological and psychiatric diseases including Alzheimer's, parasympathetic palsy, Huntington's disease, myotonic dystrophy, anorexia and depression; cardiovascular diseases including congestive heart failure, Hodgkin's disease and myocardial infarction; respiratory diseases including asthma, chronic obstructive pulmonary disease, cystic fibrosis and adult respiratory distress syndrome; liver diseases including hypercholesterolaemia, cirrhosis, viral and nonviral hepatitis, Type II diabetes mellitus, and impaired glucose tolerance; renal diseases including renal failure, acute tubular necrosis and glomerulonephritis; skeletal muscle diseases including Eulenburg's disease, hypoglycaemia and obesity; gastrointestinal diseases including myotonia congenita and intestinal obstruction; lymphomas including lymphoblastic leukaemia; diseases of placenta including choriocarcinoma; diseases of testes including testicular cancer; male reproductive diseases including low testosterone and male infertility; and disease of pancreas including diabetic ketoacidosis, Type 1 and 2 diabetes and obesity. The present sequence is a polypeptide of the invention.

Query Match	48.08	Score 47	DB 23	Length 394
Best Local Similarity	45.08	Pred. NO. 24		
Matches 9, Conservative	6	Mismatches 5	Indels 0	Gaps 0

```
QY      1 ADLSGIASGGRLYISRMGK 20
          |||::|: ||:::|
Db     317 ADLTGISPPNLYLSKIHK 336
```

Search completed: July 11, 2003, 11:54:22  
Job time : 12.4761 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:49:36 ; Search time 3.57588 Seconds

(without alignments)  
164.563 Million cell updates/sec

Title: US-10-091-442-6

Sequence: 1 ADLGIASGRLYSRMWK 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Database : Listing first 45 summaries

Issued Patents\_Aa: \*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep: \*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep: \*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep: \*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep: \*  
5: /cgn2\_6/ptodata/1/1aa/PC70S.COMB.pep: \*  
6: /cgn2\_6/ptodata/1/1aa/Backfile1.pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	98	100.0	20	2	US-08-611-977-6	Sequence 6, Appl1
2	92	93.9	20	2	US-08-472-659-6	Sequence 6, Appl1
3	92	93.9	20	2	US-08-474-661-6	Sequence 6, Appl1
4	90	91.8	380	2	US-08-472-659-34	Sequence 34, Appl1
5	90	91.8	380	2	US-08-474-661-34	Sequence 34, Appl1
6	90	91.8	380	2	US-08-611-977-34	Sequence 34, Appl1
7	46	46.9	327	4	US-09-583-492-10	Sequence 10, Appl1
8	43	43.9	375	1	US-08-121-714-8	Sequence 8, Appl1
9	43	43.9	375	1	US-08-477-108A-8	Sequence 8, Appl1
10	43	43.9	375	2	US-08-477-112-8	Sequence 8, Appl1
11	43	43.9	375	5	PCT-US93-08322-8	Sequence 8, Appl1
12	43	43.9	391	4	US-09-123-912-110	Sequence 110, Appl1
13	43	43.9	391	4	US-09-643-597-110	Sequence 110, Appl1
14	43	43.9	392	4	US-09-026-408-2	Sequence 2, Appl1
15	43	43.9	400	4	US-09-123-912-112	Sequence 112, Appl1
16	43	43.9	400	4	US-09-643-597-112	Sequence 112, Appl1
17	43	43.9	405	4	US-09-026-408-13	Sequence 13, Appl1
18	43	43.9	406	1	US-08-434-881-2	Sequence 2, Appl1
19	43	43.9	406	3	US-08-977-771-2	Sequence 2, Appl1
20	43	43.9	406	4	US-09-361-773-2	Sequence 2, Appl1
21	42	42.9	379	1	US-08-121-714-4	Sequence 4, Appl1
22	42	42.9	379	1	US-08-477-108A-4	Sequence 4, Appl1
23	42	42.9	379	5	US-08-477-112-4	Sequence 4, Appl1
24	42	42.9	379	5	PCT-US93-08322-4	Sequence 4, Appl1
25	41.5	42.3	331	4	US-09-199-637A-333	Sequence 333, Appl1
26	41	41.8	245	4	US-09-342-084-4	Sequence 4, Appl1
27	39	39.8	65	4	US-09-579-236-5	Sequence 5, Appl1

28	39	39.8	125	4	US-09-579-236-2	Sequence 2, Appl1
29	39	39.8	328	1	US-08-423-691-2	Sequence 2, Appl1
30	39	39.8	328	1	US-08-080-386-4	Sequence 4, Appl1
31	39	39.8	328	2	US-08-390-000A-4	Sequence 4, Appl1
32	39	39.8	328	3	US-08-867-260-2	Sequence 2, Appl1
33	39	39.8	328	3	US-09-501-192-9	Sequence 9, Appl1
34	39	39.8	465	3	US-08-948-997-6	Sequence 6, Appl1
35	39	39.8	465	4	US-09-348-817A-6	Sequence 6, Appl1
36	39	39.8	489	2	US-08-752-307B-3	Sequence 3, Appl1
37	39	39.8	489	4	US-09-330-317B-18	Sequence 18, Appl1
38	39	39.8	489	4	US-09-707-802-3	Sequence 3, Appl1
39	39	39.8	489	4	US-09-991-326-3	Sequence 3, Appl1
40	39	39.8	489	4	US-09-808-589A-18	Sequence 18, Appl1
41	39	39.8	506	4	US-09-360-237-59	Sequence 59, Appl1
42	39	39.8	530	2	US-08-752-307B-2	Sequence 2, Appl1
43	39	39.8	530	4	US-09-707-802-2	Sequence 2, Appl1
44	39	39.8	530	4	US-09-991-326-2	Sequence 2, Appl1
45	39	39.8	625	2	US-08-532-547-7	Sequence 7, Appl1

#### ALIGNMENTS

RESULT 1  
US-08-611-977-6  
Sequence 6, Application US/08611977  
Patent No. 5972886  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSUROUOKA, No. 5972886uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. 5972886uhlro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAUCHI, Kozo  
APPLICANT: YAMAGUCHI, No. 5972886omi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/611,977  
FILING DATE: 06-MAR-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm R.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-204  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:

LENGTH: 20 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-611-977-6

Query Match 100.0%; Score 98; DB 2; Length 20;  
Best Local Similarity 100.0%; Pred. No. 2e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ADLSGIASGRLYISRMXGK 20  
Db 1 ADLSGIASGRLYISRMXGK 20

## RESULT 2

US-08-472-659-6  
Sequence 6, Application US/08472659  
Patent No. 5831030

## GENERAL INFORMATION:

APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IMASA, Fuyuki  
APPLICANT: TSUROOKA, No. 5831030uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. 5831030uhiro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, No. 5831030oml  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/472,659  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993

## ATTORNEY/AGENT INFORMATION:

NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET INFORMATION: 001560-248  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-6620  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-472-659-6

Query Match 93.9%; Score 92; DB 2; Length 20;  
Best Local Similarity 95.0%; Pred. No. 2.1e-09;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ADLSGIASGRLYISRMXGK 20  
Db 1 ADLSGIASGRLYISRMXGK 20

## RESULT 3

US-08-474-661-6  
Sequence 6, Application US/08474661  
Patent No. 5874253

## GENERAL INFORMATION:

APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IMASA, Fuyuki  
APPLICANT: TSUROOKA, No. 5874253uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. 5874253uhiro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, No. 5874253oml  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/474,661  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993

## ATTORNEY/AGENT INFORMATION:

NAME: REA, TERESA STANER  
REGISTRATION NUMBER: 30,427  
REFERENCE/DOCKET NUMBER: 001560-204  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-6620  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-474-661-6

Query Match 93.9%; Score 92; DB 2; Length 20;  
Best Local Similarity 95.0%; Pred. No. 2.1e-09;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ADLSGIASGRLYISRMXGK 20  
Db 1 ADLSGIASGRLYISRMXGK 20



RESULT 4  
US-08-472-659-34  
Sequence 34, Application US/08472659  
Patent No. 5831030  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSUROUOKA, No. 5831030uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. 5831030uh1ro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, No. 5831030om1  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/472,659  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-248  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-6620  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-472-659-34

Query Match 91.8%; Score 90; DB 2; Length 380;  
Best Local Similarity 95.0%; Pred. No. 1.6e-07;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ADLSGIASGRLYISRMWK 20  
Db 305 ADLSGIASGRLYISRMWK 324

RESULT 5  
US-08-474-661-34  
Sequence 34, Application US/08474661  
Patent No. 5874253  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki

APPLICANT: TSUROKA, No. 5874253uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. 5874253uh1ro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, No. 5874253om1  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/474,661  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: REA, TERESA STANER  
REGISTRATION NUMBER: 30,427  
REFERENCE/DOCKET NUMBER: 001560-204  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-6620  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-474-661-34

Query Match 91.8%; Score 90; DB 2; Length 380;  
Best Local Similarity 95.0%; Pred. No. 1.6e-07;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ADLSGIASGRLYISRMWK 20  
Db 305 ADLSGIASGRLYISRMWK 324

RESULT 6  
US-08-611-977-34  
Sequence 34, Application US/08611977  
Patent No. 5972886  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSUROUOKA, No. 5972886uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. 5972886uh1ro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, No. 5972886om1  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/611,977  
FILING DATE: 06-MAR-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-204  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-611-977-34

Query Match 91.8%; Score 90; DB 2; Length 380;  
Best Local Similarity 95.0%; Pred. No. 1,6e-07;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ADLSGASGRLYISRMK 20  
Db 305 ADLSGASGRLYISRMHK 324

RESULT 7  
US-09-583-492-10  
Sequence 10, Application US/09583492  
Patent No. 6407222  
GENERAL INFORMATION:  
APPLICANT: Cul, Jisong  
APPLICANT: Lo, Jane-Ling  
APPLICANT: Mount, George R.  
TITLE OF INVENTION: DOG GONADOTROPIN RELEASING HORMONE  
FILE REFERENCE: 20140  
CURRENT APPLICATION NUMBER: US/09/583,492  
EARLIER FILING DATE: 2000-05-31  
EARLIER APPLICATION NUMBER: 60/138,448  
NUMBER OF SEQ ID NOS: 10  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 10  
LENGTH: 327  
TYPE: PRT  
ORGANISM: canis  
US-09-583-492-10

Query Match 46.9%; Score 46; DB 4; Length 327;

Best Local Similarity 62.5%; Pred. No. 4.2;  
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 LSGIASGRLYISRM 18  
Db 165 LSGIAGPOLYIFRM 180

RESULT 8  
US-08-121-714-8  
Sequence 8, Application US/08121714  
Patent No. 5470970  
GENERAL INFORMATION:  
APPLICANT: Sager, Ruth  
TITLE OF INVENTION: MASPIN, A NOVEL SERPIN WITH  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM PS/2 Model 502 or 555X  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: Wordperfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/121,714  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/938,823  
FILING DATE: 09/01/92  
APPLICATION NUMBER: 07/844,296  
FILING DATE: 02/28/92  
APPLICATION NUMBER: 07/662,216  
FILING DATE: 02/28/91  
ATTORNEY/AGENT INFORMATION:  
NAME: Fraser, Janis K.  
REGISTRATION NUMBER: 34,819  
REFERENCE/DOCKET NUMBER: 00530/072001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 375  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-08-121-714-8

Query Match 43.9%; Score 43; DB 1; Length 375;  
Best Local Similarity 40.0%; Pred. No. 16;  
Matches 8; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

QY 1 ADLSGASGRLYISRMK 20  
Db 302 ADLSGMSGARDLFVSKTIHK 321

RESULT 9  
US-08-477-108A-8  
Sequence 8, Application US/08477108A  
Patent No. 5801001  
GENERAL INFORMATION:  
APPLICANT: Sager, Ruth  
APPLICANT: Zou, Zhiqiang  
APPLICANT: Anisowicz, Anthony

;; TITLE OF INVENTION: MASPIN, A NOVEL SERPIN WITH  
;; TITLE OF INVENTION: TUMOR SUPPRESSING ACTIVITY  
;; NUMBER OF SEQUENCES: 11  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Fish & Richardson P.C.  
;; STREET: 225 Franklin Street  
;; CITY: Boston  
;; STATE: Massachusetts  
;; COUNTRY: U.S.A.  
;; ZIP: 02110-2804  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
;; COMPUTER: IBM PS/2 Model 502 or 55SX  
;; OPERATING SYSTEM: MS-DOS (Version 5.0)  
;; SOFTWARE: WordPerfect (Version 5.1)  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/477,108A  
;; FILING DATE: June 7, 1995  
;; CLASSIFICATION: 536  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/121,714  
;; FILING DATE: 09/01/93  
;; APPLICATION NUMBER: 07/938,823  
;; FILING DATE: 09/01/92  
;; APPLICATION NUMBER: 07/844,296  
;; FILING DATE: 02/28/92  
;; APPLICATION NUMBER: 07/662,216  
;; FILING DATE: 02/28/91  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Fraser, Janis K.  
;; REGISTRATION NUMBER: 34,819  
;; REFERENCE/DOCKET NUMBER: 06570/002002  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (617) 542-5070  
;; TELEFAX: (617) 542-8906  
;;  
;; INFORMATION FOR SEQ ID NO: 8:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 375  
;; TYPE: amino acid  
;; STRANDEDNESS:  
;; TOPOLOGY: linear  
;;  
;; US-08-477-108A-8

Query Match 43.9%; Score 43; DB 1; Length 375;  
Best Local Similarity 40.0%; Pred. No. 16;  
Matches 8; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ADLGGIAGGRLYISRMGK 20  
|||||:|:|:|:|  
Db 302 ADLGGMGGARDLFFVSKIHK 321

RESULT 10  
US-08-477-112-8  
; Sequence 8, Application US/08477112  
; Patent No. 5905023  
; GENERAL INFORMATION:  
; APPLICANT: Sager, Ruth  
; TITLE OF INVENTION: MASPIN, A NOVEL SERPIN WITH  
; TITLE OF INVENTION: TUMOR SUPPRESSING ACTIVITY  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; COMPUTER: IBM PS/2 Model 502 or 55SX  
; OPERATING SYSTEM: MS-DOS (Version 5.0)

;; SOFTWARE: WordPerfect (Version 5.1)  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/477,112  
;; FILING DATE:  
;; CLASSIFICATION: 536  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/121,714  
;; FILING DATE: 09/01/93  
;; APPLICATION NUMBER: 07/938,823  
;; FILING DATE: 09/01/92  
;; APPLICATION NUMBER: 07/844,296  
;; FILING DATE: 02/28/92  
;; APPLICATION NUMBER: 07/662,216  
;; FILING DATE: 02/28/91  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Fraser, Janis K.  
;; REGISTRATION NUMBER: 34,819  
;; REFERENCE/DOCKET NUMBER: 06570/002003  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (617) 542-5070  
;; TELEFAX: (617) 542-8906  
;;  
;; INFORMATION FOR SEQ ID NO: 8:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 375  
;; TYPE: amino acid  
;; STRANDEDNESS:  
;; TOPOLOGY: linear  
;;  
;; US-08-477-112-8

Query Match 43.9%; Score 43; DB 2; Length 375;  
Best Local Similarity 40.0%; Pred. No. 16;  
Matches 8; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ADLGGIAGGRLYISRMGK 20  
|||||:|:|:|:|  
Db 302 ADLGGMGGARDLFFVSKIHK 321

RESULT 11  
PCT-US93-08322-8  
; Sequence 8, Application PC/TUS9308322  
; GENERAL INFORMATION:  
; APPLICANT: Sager, Ruth  
; TITLE OF INVENTION: MASPIN, A NOVEL SERPIN WITH TUMOR SUPPRESSING ACTIVITY  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; COMPUTER: IBM PS/2 Model 502 or 55SX  
; OPERATING SYSTEM: MS-DOS (Version 5.0)  
; SOFTWARE: WordPerfect (Version 5.1)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US93/08322  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/938,823  
; FILING DATE: 09/01/92  
; APPLICATION NUMBER: 07/844,296  
; FILING DATE: 02/28/92  
; APPLICATION NUMBER: 07/662,216  
; FILING DATE: 02/28/91  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fraser, Janis K.  
; REGISTRATION NUMBER: 34,819  
; REFERENCE/DOCKET NUMBER: 00530/072001

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 375
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
PCT-US93-08322-8

Query Match          43.9%; Score 43; DB 5; Length 375;
Best Local Similarity 40.0%; Pred. No. 16;
Matches 8; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

QY      1 ADLSGASGRLYISRMKG 20
Db      302 ADLSGMSGARDLPVSKITIK 321

RESULT 12
US-09-123-912-110
; Sequence 110, Application US/09123912A
; Patent No. 6312695
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Wang, Tonglong
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY OF LUNG CANCER
; FILE REFERENCE: 210121.455c1
; CURRENT APPLICATION NUMBER: US/09/123,912A
; CURRENT FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: 09/040,802
; PRIOR FILING DATE: 1998-03-18
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 110
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-123-912-110

Query Match          43.9%; Score 43; DB 4; Length 391;
Best Local Similarity 44.4%; Pred. No. 17;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY      1 ADLSGASGRLYISRM 18
Db      314 ADYSGMSSGSLVAQKFL 331

RESULT 13
US-09-643-597-110
; Sequence 110, Application US/09643597
; Patent No. 6426072
; GENERAL INFORMATION:
; APPLICANT: Wang, Tonglong
; APPLICANT: Fan, Liqun
; APPLICANT: Kalos, Michael D.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Hosken, Nancy
; APPLICANT: Fanger, Gary R.
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aljun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Henderson, Robert A.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.455c11
; CURRENT APPLICATION NUMBER: US/09/643,597
; CURRENT FILING DATE: 2000-08-21
; NUMBER OF SEQ ID NOS: 369
```

```

; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 110
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-643-597-110

Query Match          43.9%; Score 43; DB 4; Length 391;
Best Local Similarity 44.4%; Pred. No. 17;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY      1 ADLSGASGRLYISRM 18
Db      314 ADYSGMSSGSLVAQKFL 331

RESULT 14
US-09-026-408-2
; Sequence 2, Application US/09026408
; Patent No. 630338
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: PANCREAS-DERIVED PLASMINOGEN ACTIVATOR
; TITLE OF INVENTION: INHIBITOR
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/026,408
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/934,011
; FILING DATE: 13-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,056
; FILING DATE: 16-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0300002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 392 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-026-408-2

Query Match          43.9%; Score 43; DB 4; Length 392;
Best Local Similarity 42.1%; Pred. No. 17;
Matches 8; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY      2 DLSGASGRLYISRMKG 20
Db      316 DLSGITDSSEVYVSOYTK 334

RESULT 15
US-09-123-912-112
```

Fri Jul 11 15:11:11 2003

us-10-091-442-6.ra1

Page 7

```

: Sequence 112, Application US/09123912A
: Patent No. 6312695
: GENERAL INFORMATION:
: APPLICANT: Reed, Steven G.
: APPLICANT: Wang, Tongtong
: TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY OF LUNG CANCER
: FILE REFERENCE: 210121.455C1
: CURRENT APPLICATION NUMBER: US/09/123,912A
: CURRENT FILING DATE: 1998-07-27
: PRIORITY APPLICATION NUMBER: 09/040,802
: PRIOR FILING DATE: 1998-03-18
: NUMBER OF SEQ ID NOS: 114
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 112
: LENGTH: 400
: TYPE: prt
: ORGANISM: Homo sapiens
: US-09-123-912-112

```

Query Match	43.9%	Score 43	DB 4	Length 400
Best Local Similarity	44.4%	Pred. No. 17		
Matches	8	Conservative	4	Mismatches 6
				Indels 0
				Gaps 0
QY	1	ADLSGIASGRLIISRM	18	
		-----	:	
Db	323	ADYSGMSSGGLYAKFL	340	

Search completed: July 11, 2003, 12:02:09  
Job time : 4.57588 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: July 11, 2003, 11:59:17 ; Search time 5.86279 Seconds  
(without alignments)  
397.148 Million cell updates/sec

Title: US-10-091-442-6  
Perfect score: 98  
Sequence: 1 ADLSGIASGGRLYSRMWGK 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 445758 seqs, 116419773 residues

Total number of hits satisfying chosen parameters: 445758

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

Published\_Applications\_AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_NEM\_PUB pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PC1\_NEM\_PUB pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEM\_PUB pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/PC1US\_PUBCOMB pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEM\_PUB pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEM\_PUB pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEM\_PUB pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEM\_PUB pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	98	100.0	20	9 US-10-091-442-6	Sequence 6, Appli
2	98	100.0	20	10 US-10-140-719-6	Sequence 6, Appli
3	90	91.8	380	9 US-10-091-442-34	Sequence 34, Appli
4	90	91.8	380	10 US-09-110-219-34	Sequence 34, Appli
5	47	48.0	395	9 US-09-823-187-2	Sequence 2, Appli
6	45	45.9	307	9 US-09-738-626-6331	Sequence 6331, Ap
7	45	45.9	379	9 US-10-097-340-139	Sequence 139, App
8	44	44.9	377	9 US-10-165-605A-27	Sequence 27, Appli
9	44	44.9	377	10 US-09-910-430-37	Sequence 27, Appli
10	43	43.9	391	9 US-10-157-031-355	Sequence 355, App
11	43	43.9	391	9 US-09-466-396A-110	Sequence 110, App
12	43	43.9	391	10 US-09-735-705-110	Sequence 110, App
13	43	43.9	391	10 US-09-850-716A-110	Sequence 110, App
14	43	43.9	391	10 US-09-897-778-110	Sequence 110, App
15	43	43.9	392	10 US-09-903-684-2	Sequence 2, Appli
16	43	43.9	400	9 US-09-466-396A-112	Sequence 112, App
17	43	43.9	400	10 US-09-735-705-112	Sequence 112, App
18	43	43.9	400	10 US-09-850-716A-112	Sequence 112, App
19	43	43.9	400	10 US-09-897-778-112	Sequence 112, App

20	43	43.9	405	10 US-09-902-684-13	Sequence 13, Appli
21	42	42.9	437	10 US-09-925-301-1200	Sequence 1200, Ap
22	41.5	42.3	331	9 US-09-975-719-333	Sequence 333, App
23	41	41.8	143	9 US-10-091-504-679	Sequence 679, App
24	41	41.8	143	10 US-09-764-869-679	Sequence 679, App
25	41	41.8	217	10 US-09-729-674-134	Sequence 134, App
26	41	41.8	340	10 US-09-764-898-169	Sequence 169, App
27	41	41.8	379	9 US-10-012-896-946	Sequence 946, App
28	41	41.8	392	9 US-09-895-814-946	Sequence 946, App
29	41	41.8	392	9 US-10-198-070-62	Sequence 62, Appli
30	41	41.8	392	9 US-10-198-070-67	Sequence 67, Appli
31	41	41.8	392	9 US-10-198-070-70	Sequence 70, Appli
32	41	41.8	402	9 US-10-135-629-4	Sequence 4, Appli
33	41	41.8	413	9 US-09-993-180-3	Sequence 3, Appli
34	41	41.8	423	9 US-10-135-629-3	Sequence 3, Appli
35	41	41.8	423	9 US-09-919-039-19	Sequence 19, Appli
36	41	41.8	423	9 US-09-978-418-14	Sequence 14, Appli
37	41	41.8	433	9 US-10-205-823-361	Sequence 361, App
38	41	41.8	433	9 US-10-177-293-421	Sequence 421, App
39	41	41.8	617	9 US-10-012-896-947	Sequence 947, App
40	41	41.8	617	9 US-09-895-814-947	Sequence 947, App
41	41	41.8	91	10 US-09-925-302-886	Sequence 886, App
42	40	40.8	65	10 US-09-754-016-5	Sequence 5, Appli
43	39	39.8	125	10 US-09-754-016-2	Sequence 2, Appli
44	39	39.8	328	9 US-10-225-567A-136	Sequence 136, App
45	39	39.8	328	9 US-10-225-567A-136	Sequence 136, App

## ALIGNMENTS

### RESULT 1

US-10-091-442-6  
Sequence 6, Application US/10091442

Patent No. US20020164711A1  
GENERAL INFORMATION:

APPLICANT: TSUJIMOTO, Masafumi  
IWASA, Fuyuki

TSURUOKA, No. US20020164711A1uo  
NAKAZATO, Hiroshi

MURA, Kenju  
ISHIDA, No. US20020164711A1uhfro

KURIHARA, Tatsuya  
YAMAICHI, Koza

YAMAGUCHI, No. US20020164711A1omf  
TITLE OF INVENTION: MEKANARYOCYTE DIFFERENTIATION FACTOR

NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404

CITY: Alexandria  
STATE: Virginia

COUNTRY: United States  
ZIP: 22313-1404

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/091,442

FILING DATE: 07-Mar-2002  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719

FILING DATE: 08-AUG-1998  
APPLICATION NUMBER: US 08/474,661

FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/091,028

FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-212305

FILING DATE: 17-JUL-1992  
APPLICATION NUMBER: JP 6-067339

FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
US-10-091-442-6

Query Match 100.0%; Score 98; DB 9; Length 20;  
Best Local Similarity 100.0%; Pred. No. 4.2e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLSGIASGGRLYISRMGK 20  
DB 1 ADLSGIASGGRLYISRMGK 20

RESULT 2  
US-09-140-719-6  
Sequence 6, Application US/09140719  
Patent No. US20010026931A1  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSUROUOKA, No. US20010026931A1uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MURA, Kenju  
APPLICANT: ISHIDA, No. US20010026931A1uhltro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAGUCHI, Kozo  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.

REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-140-719-6

Query Match 100.0%; Score 98; DB 10; Length 20;  
Best Local Similarity 100.0%; Pred. No. 4.2e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLSGIASGGRLYISRMGK 20  
DB 1 ADLSGIASGGRLYISRMGK 20

RESULT 3  
US-10-091-442-34  
Sequence 34, Application US/10091442  
Patent No. US20020164711A1  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSUROUOKA, No. US20020164711A1uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MURA, Kenju  
APPLICANT: ISHIDA, No. US20020164711A1uhltro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAGUCHI, Kozo  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/091,442  
FILING DATE: 07-MAR-2002  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: <Unknown>  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021



INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids,  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 34:  
US-10-091-442-34

Query Match 91.8%; Score 90; DB 9; Length 380;  
Best Local Similarity 95.0%; Pred. NO. 2e-06;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ADLSGIASGRLYISRMKG 20  
Db 305 ADLSGIASGRLYISRMKG 324

RESULT 4  
US-09-140-719-34  
Sequence 34, Application US/09140719  
Patent No. US20010026931A1  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masaumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSURUOKA, No. US20010026931A1uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MURA, Kenju  
APPLICANT: ISHIDA, No. US20010026931A1uhlro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAUCHI, Kozo  
APPLICANT: YAMAGUCHI, No. US20010026931A1oml  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid

TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-140-719-34

Query Match 91.8%; Score 90; DB 10; Length 380;  
Best Local Similarity 95.0%; Pred. NO. 2e-06;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ADLSGIASGRLYISRMKG 20  
Db 305 ADLSGIASGRLYISRMKG 324

RESULT 5  
US-09-823-187-2  
Sequence 2, Application US/09823187  
Publication No. US20030096952A1  
GENERAL INFORMATION:  
APPLICANT: Burgess, Catherine  
APPLICANT: Gusev, Vladimir Y  
APPLICANT: Liu, Xiaohong  
APPLICANT: Majumder, Kumud  
APPLICANT: Padigaru, Muralidhar  
APPLICANT: Patturajan, Meera  
APPLICANT: Shimkets, Richard A  
APPLICANT: Spaderna, Steven K  
APPLICANT: Spytek, Kimberly  
APPLICANT: Taupier, Raymond J  
TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME  
FILE REFERENCE: 15966-745  
CURRENT APPLICATION NUMBER: US/09/823,187  
CURRENT FILING DATE: 2001-03-29  
PRIOR APPLICATION NUMBER: 60/193,339  
PRIOR FILING DATE: 2000-03-30  
PRIOR APPLICATION NUMBER: 60/195,343  
PRIOR FILING DATE: 2000-04-05  
PRIOR APPLICATION NUMBER: 60/195,088  
PRIOR FILING DATE: 2000-04-06  
PRIOR APPLICATION NUMBER: 60/195,005  
PRIOR FILING DATE: 2000-04-06  
PRIOR APPLICATION NUMBER: 60/195,792  
PRIOR FILING DATE: 2000-04-10  
PRIOR APPLICATION NUMBER: 60/196,556  
PRIOR FILING DATE: 2000-04-11  
PRIOR APPLICATION NUMBER: 60/197,081  
PRIOR FILING DATE: 2000-04-13  
PRIOR APPLICATION NUMBER: 60/197,525  
PRIOR FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/197,087  
PRIOR FILING DATE: 2000-04-14  
NUMBER OF SEQ ID NOS: 103  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 395  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-823-187-2

Query Match 48.0%; Score 47; DB 9; Length 395;  
Best Local Similarity 45.0%; Pred. NO. 14;  
Matches 9; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 1 ADLSGIASGRLYISRMKG 20  
Db 318 ADLTGISPNLVLSKITHK 337

RESULT 6  
US-09-738-626-6331  
Sequence 6331, Application US/09738626  
Publication No. US20020197605A1

GENERAL INFORMATION:  
APPLICANT: NAKAGAWA, SATOSHI  
APPLICANT: MIZOGUCHI, HIROSHI  
APPLICANT: ANDO, SEIKO  
APPLICANT: HAYASHI, MIKIRO  
APPLICANT: YOKOI, KEIKO  
APPLICANT: YOKOI, HARUHIKO  
APPLICANT: TATEISHI, NAOKO  
APPLICANT: SENOH, AKIHIRO  
APPLICANT: IKEDA, MASATO  
APPLICANT: OZAKI, AKIO  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-125  
CURRENT APPLICATION NUMBER: US/09/738,626  
CURRENT FILING DATE: 2000-12-18  
PRIOR APPLICATION NUMBER: JP 99/377484  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: JP 00/159162  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: JP 00/280988  
PRIOR FILING DATE: 2000-08-03  
NUMBER OF SEQ ID NOS: 7059  
SOFTWARE: PatentIn ver. 3.0  
SEQ ID NO 6331  
LENGTH: 307  
TYPE: PRT  
ORGANISM: Corynebacterium glutamicum  
US-09-738-626-6331

Query Match 45.9%; Score 45; DB 9; Length 307;  
Best Local Similarity 50.0%; Pred. No. 22;  
Matches 9; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 3 LSGIAGSGRLYISRMKG 20  
DB 215 LSGIVTGVMLAVARVNGE 232

RESULT 7  
US-10-097-340-139  
Sequence 139, Application US/10097340  
Publication No. US20030087250A1  
GENERAL INFORMATION:  
APPLICANT: JOHN MONAHAN  
APPLICANT: Manjula GANNAVARAPU  
APPLICANT: Sebastian HOERSCH  
APPLICANT: Shubhangi KAMATKAR  
APPLICANT: Steve G. KOVATS  
APPLICANT: Rachel E. MEYERS  
APPLICANT: Michael MORRISSEY  
APPLICANT: Peter OLANDT  
APPLICANT: Ami SEN  
APPLICANT: Peter VEIBY  
APPLICANT: Gordon B. MILLS  
APPLICANT: Robert C. BAST, Jr.  
APPLICANT: Karen LU  
APPLICANT: Rosemarie SCHMANDT  
APPLICANT: Xumei ZHAO  
APPLICANT: Karen GLATT  
TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,  
Title Of Invention: Assessment, Prevention, and Therapy Of Ovarian Cancer  
FILE REFERENCE: MRI-030  
CURRENT APPLICATION NUMBER: US/10/097,340  
CURRENT FILING DATE: 2002-03-14  
PRIOR APPLICATION NUMBER: 60/276,025  
PRIOR FILING DATE: 2001-03-14  
PRIOR APPLICATION NUMBER: 60/325,149  
PRIOR FILING DATE: 2001-09-26  
PRIOR APPLICATION NUMBER: 60/276,026  
PRIOR FILING DATE: 2001-03-14  
PRIOR APPLICATION NUMBER: 60/324,967  
PRIOR FILING DATE: 2001/09/26  
PRIOR APPLICATION NUMBER: 60/311,732

PRIOR FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: 60/325,102  
PRIOR FILING DATE: 2001-09-26  
PRIOR APPLICATION NUMBER: 60/323,580  
PRIOR FILING DATE: 2001-09-19  
NUMBER OF SEQ ID NOS: 363  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 139  
LENGTH: 379  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-097-340-139

Query Match 45.9%; Score 45; DB 9; Length 379;  
Best Local Similarity 45.0%; Pred. No. 28;  
Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 1 ADLSGIAGSGRLYISRMKG 20  
DB 304 ADLSGISNHSNIQVSEVHK 323

RESULT 8  
US-10-165-605A-27  
Sequence 27, Application US/10165605A  
Publication No. US20030086937A1  
GENERAL INFORMATION:  
APPLICANT: Godfroid, Edmond  
APPLICANT: Bollen, Alex  
APPLICANT: Leboulle, Gerard  
TITLE OF INVENTION: IDENTIFICATION AND MOLECULAR CHARACTERIZATION OF  
TITLE OF INVENTION: PROTEINS, EXPRESSED IN THE IXODS RICINUS SALIVARY  
FILE REFERENCE: VANM229.001CP2  
CURRENT APPLICATION NUMBER: US/10/165,605A  
CURRENT FILING DATE: 2002-06-07  
PRIOR APPLICATION NUMBER: 09/910,430  
PRIOR FILING DATE: 2001-07-19  
PRIOR APPLICATION NUMBER: PCT/BE00/00061  
PRIOR FILING DATE: 2000-06-06  
PRIOR APPLICATION NUMBER: GB9913425.6  
PRIOR FILING DATE: 1998-06-09  
NUMBER OF SEQ ID NOS: 34  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 27  
LENGTH: 377  
TYPE: PRT  
ORGANISM: Ixodes ricinus  
US-10-165-605A-27

Query Match 44.9%; Score 44; DB 9; Length 377;  
Best Local Similarity 50.0%; Pred. No. 40;  
Matches 10; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 ADLSGIAGSGRLYISRMKG 20  
DB 298 ADLSGISAGNLCASDVHK 317

RESULT 9  
US-09-910-430-27  
Sequence 27, Application US/09910430  
Patent No. US20020127235A1  
GENERAL INFORMATION:  
APPLICANT: Godfroid, Edmond  
APPLICANT: Bollen, Alex  
APPLICANT: Leboulle, Gerard  
TITLE OF INVENTION: IDENTIFICATION AND MOLECULAR CHARACTERIZATION OF  
TITLE OF INVENTION: PROTEINS, EXPRESSED IN THE IXODS RICINUS SALIVARY  
FILE REFERENCE: VANM229.001CP1  
CURRENT APPLICATION NUMBER: US/09/910,430  
CURRENT FILING DATE: 2001-07-19

PRIOR APPLICATION NUMBER: PCT/BE00/00061  
PRIOR FILING DATE: 2000-06-06  
PRIOR APPLICATION NUMBER: GB 9913425.6  
PRIOR FILING DATE: 1999-06-09  
NUMBER OF SEQ ID NOS: 34  
SOFTWARE: FASTSEQ for Windows Version 3.0  
SEQ ID NO: 27  
LENGTH: 377  
TYPE: PRT  
ORGANISM: Ixodes ricinus  
US-09-910-430-27

Query Match 44.9%; Score 44; DB 10; Length 377;  
Best Local Similarity 50.0%; Pred. No. 40;  
Matches 10; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 1 ADLSGIASGRLYISRMCK 20  
DB 298 ADLSGIASGRLYISRMCK 317

RESULT 10  
US-10-157-031-355  
Sequence 355, Application US/10157031  
Publication No. US20030108890A1  
GENERAL INFORMATION:  
APPLICANT: Baranova, A. V.  
APPLICANT: Yankovsky, N. K.  
APPLICANT: Kozlov, A. P.  
APPLICANT: Lobashov, A. V.  
APPLICANT: Krukovskaya, L. L.  
TITLE OF INVENTION: In silico screening for phenotype-associated expressed sequences  
FILE REFERENCE: 2760-103  
CURRENT APPLICATION NUMBER: US/10/157,031  
CURRENT FILING DATE: 2002-05-30  
NUMBER OF SEQ ID NOS: 415  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO: 355  
LENGTH: 391  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-157-031-355

Query Match 43.9%; Score 43; DB 9; Length 391;  
Best Local Similarity 44.4%; Pred. No. 61;  
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 1 ADLSGIASGRLYISRM 18  
DB 314 ADLSGIASGRLYISRM 331

RESULT 11  
US-09-466-396A-110  
Sequence 110, Application US/09466396A  
Publication No. US20030119763A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Tonglong  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY AND  
FILE REFERENCE: 210121.455C4  
CURRENT APPLICATION NUMBER: US/09/466,396A  
CURRENT FILING DATE: 1999-12-17  
NUMBER OF SEQ ID NOS: 224  
SOFTWARE: FASTSEQ for Windows Version 3.0  
SEQ ID NO: 110  
LENGTH: 391  
TYPE: PRT  
ORGANISM: Homo sapien  
US-09-466-396A-110

Query Match 43.9%; Score 43; DB 9; Length 391;  
Best Local Similarity 44.4%; Pred. No. 61;

Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;  
OY 1 ADLSGIASGRLYISRM 18  
DB 314 ADLSGIASGRLYISRM 331

RESULT 12  
US-09-735-705-110  
Sequence 110, Application US/09735705  
Patent No. US20020052329A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Tonglong  
APPLICANT: Fan, Liqun  
APPLICANT: Kalos, Michael D.  
APPLICANT: Bangur, Chaitanya S.  
APPLICANT: Hosken, Nancy  
APPLICANT: Fanger, Gary R.  
APPLICANT: Li, Samuel X.  
APPLICANT: Wang, Aijun  
APPLICANT: Skelky, Yasir A.W.  
APPLICANT: Henderson, Robert A.  
APPLICANT: McNeill, Patricia D.  
APPLICANT: Fanger, Neil  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
FILE REFERENCE: 210121.455C14  
CURRENT APPLICATION NUMBER: US/09/735,705  
CURRENT FILING DATE: 2000-12-12  
NUMBER OF SEQ ID NOS: 419  
SOFTWARE: FASTSEQ for Windows Version 3.0  
SEQ ID NO: 110  
LENGTH: 391  
TYPE: PRT  
ORGANISM: Homo sapien  
US-09-735-705-110

Query Match 43.9%; Score 43; DB 10; Length 391;  
Best Local Similarity 44.4%; Pred. No. 61;  
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 1 ADLSGIASGRLYISRM 18  
DB 314 ADLSGIASGRLYISRM 331

RESULT 13  
US-09-850-716A-110  
Sequence 110, Application US/09850716A  
Patent No. US20020115139A1  
GENERAL INFORMATION:  
APPLICANT: Kalos, Michael D.  
APPLICANT: McNeill, Patricia D.  
APPLICANT: Retter, Marc W.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
FILE REFERENCE: 210121.455C15  
CURRENT APPLICATION NUMBER: US/09/850,716A  
CURRENT FILING DATE: 2001-05-07  
NUMBER OF SEQ ID NOS: 440  
SOFTWARE: FASTSEQ for Windows Version 3.0  
SEQ ID NO: 110  
LENGTH: 391  
TYPE: PRT  
ORGANISM: Homo sapien  
US-09-850-716A-110

Query Match 43.9%; Score 43; DB 10; Length 391;  
Best Local Similarity 44.4%; Pred. No. 61;  
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;  
OY 1 ADLSGIASGRLYISRM 18

Db 314 ADYSGMSSGSLVAKFL 331

RESULT 14

US-09-897-778-110

Sequence 110, Application US/09897778

Patent No. US20020147143A1

GENERAL INFORMATION:

APPLICANT: Wang, Tongtong

APPLICANT: Marnerakis, Margarita

APPLICANT: Fanger, Gary R.

APPLICANT: Vedvick, Thomas S.

APPLICANT: Carter, Darick

APPLICANT: Matanabe, Yoshihiro

APPLICANT: Henderson, Robert A.

APPLICANT: Peckham, David W.

APPLICANT: Fanger, Neil

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER

FILE REFERENCE: 210121.455C16

CURRENT APPLICATION NUMBER: US/09/897,778

CURRENT FILING DATE: 2001-06-28

NUMBER OF SEQ ID NOS: 467

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 110

LENGTH: 391

TYPE: PRT

ORGANISM: Homo sapiens

US-09-897-778-110

Query Match 43.9%; Score 43; DB 10; Length 391;

Best Local Similarity 44.4%; Pred. No. 61;

Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 1 ADLSGASGRLYISRMK 18

Db 314 ADYSGMSSGSLVAKFL 331

RESULT 15

US-09-902-684-2

Sequence 2, Application US/09902684

Patent No. US20020127640A1

GENERAL INFORMATION:

APPLICANT: Ni et al.

TITLE OF INVENTION: PANCREAS-DERIVED PLASMINOGEN ACTIVATOR

INHIBITOR

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.

STREET: 1100 NEW YORK AVENUE, SUITE 600

CITY: WASHINGTON

STATE: DC

COUNTRY: USA

ZIP: 20005-3934

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/902,684

FILING DATE: 12-Jul-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/026,408

FILING DATE: <Unknown>

APPLICATION NUMBER: US 60/024,056

FILING DATE: 16-Aug-1996

ATTORNEY/AGENT INFORMATION:

NAME: STEFFE, ERIC K.

REGISTRATION NUMBER: 36,688

REFERENCE/DOCKET NUMBER: 1488.0300002

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-371-2600

TELEFAX: 202-371-2540

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 392 amino acids

TYPE: amino acid

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-902-684-2

Query Match 43.9%; Score 43; DB 10; Length 392;

Best Local Similarity 42.1%; Pred. No. 61;

Matches 8; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 2 DLGSIASGRLYISRMK 20

Db 316 DLGSISSSEYVSVYTK 334

Search completed: July 11, 2003, 12:37:33

Job time : 5.86279 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: July 11, 2003, 11:45:06 ; Search time 4.0326 Seconds  
(without alignments)  
476.708 Million cell updates/sec

Title: US-10-091-442-6

Perfect score: 98  
Sequence: 1 ADLGIASGRLYISRMWGK 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

1: PIR\_73:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR3:\*  
5: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	54	55.1	191	2	S49162
2	48	49.0	591	2	A75474
3	46	46.9	524	2	B64302
4	45	45.9	406	2	A39339
5	45	45.9	957	2	T10633
6	44.5	45.4	434	2	C36144
7	44	44.9	205	2	D69804
8	44	44.9	222	2	C72657
9	44	44.9	453	2	AD1556
10	44	44.9	453	2	AF1198
11	44	44.9	870	2	T10634
12	43.5	44.4	1805	2	T02712
13	43	43.9	310	2	J00740
14	43	43.9	371	2	G84398
15	43	43.9	379	2	A42421
16	43	43.9	391	2	JC7118
17	43	43.9	423	2	B72528
18	43	43.9	622	2	A64494
19	43	43.9	627	2	G81719
20	43	43.9	628	2	G71565
21	42.5	43.4	203	2	D87567
22	42	42.9	126	2	G71269
23	42	42.9	311	2	D97450
24	42	42.9	311	2	AF2668
25	42	42.9	341	2	D98296
26	42	42.9	341	2	AD2887
27	42	42.9	379	2	S27383
28	42	42.9	396	2	I50707
29	42	42.9	417	2	S19724

30	42	42.9	700	2	D75508	serine/threonine p
31	42	42.9	740	2	T45091	catalase (EC 1.11.
32	42	42.9	1288	2	T09908	hypothetical prote
33	41	41.8	378	2	S38962	serpin - pig
34	41	41.8	383	2	S11433	ovalbumin - Japane
35	41	41.8	388	2	AF1907	hypothetical prote
36	41	41.8	397	2	I39184	bomapsin - human
37	41	41.8	418	2	JX0129	contrapsin precurs
38	41	41.8	433	1	ITMNC	alpha-1-antichymot
39	41	41.8	447	2	A72358	conserved hypothet
40	41	41.8	485	2	C86336	hypothetical prote
41	41	41.8	487	2	T05679	H+-transporting tw
42	41	41.8	488	2	T43789	H+-transporting tw
43	41	41.8	492	2	A31886	H+-exporting ATPas
44	41	41.8	513	2	C44138	vacuolar H+-transp
45	41	41.8	513	2	A33281	H+-transporting tw

## ALIGNMENTS

RESULT 1  
S49162  
ZG-21P protein - rat  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 16-Feb-1995 #sequence\_revision 12-May-1995 #text\_change 21-Jan-2000  
C:Accession: S49162  
R:Cronshagen, U.; Chen, C.; Kern, H.F.  
A:Submitted to the EMBL Data Library, March 1994  
A:Description: A novel protein expressed exclusively in pancreas is proposed to be a  
A:Reference number: S49162  
A:Accession: S49162  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-191 <CRO>  
A:Cross-references: EMBL:Z30585; NID:9510193; PIDN:CAA83060.1; PID:9510194  
C:Superfamily: antithrombin III

Query Match 55.1% Score 54; DB 2; Length 191;  
Best Local Similarity 57.9%; Pred. No. 0.3;  
Matches 11; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

OY 2 DLGIASGRLYISRMWGK 20  
DB 102 DLGITDSEELYSRAMQK 120

## RESULT 2

A75474  
serine proteinase, subtilase family - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000  
C:Accession: A75474

R:White, O.; Eisen, J.A.; Heldeberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.  
M.; Shen, M.; Vamathavan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.;  
Science 266, 1571-1577, 1999

A>Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MID:20036896; PMID:10567266

A:Status: preliminary  
A:Accession: A75474  
A:Molecule type: DNA

A:Residues: 1-591 <WHI>  
A:Cross-references: GB:AE001935; GB:AE000513; NID:96458517; PIDN:AAF10389.1; PID:9645

A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DR0812

A:Map position: 1  
C:Superfamily: subtilisin; subtilisin homology

Query Match 49.0% Score 48; DB 2; Length 591;  
Best Local Similarity 47.1%; Pred. No. 9.1;  
Matches 8; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

OY 3 LSGIASGRLYISRMNG 19  
 DB 255 VCGVAGSVNLYMARVLG 271

## RESULT 3

B64302  
 hypothetical protein M00018 - Methanococcus jannaschii

C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 22-Oct-1999

C:Accession: B64302

R:Smith, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, R.; Nelson, C.F.; Overbeek, R.; Kirsch, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;

Science 273, 1058-1073, 1996

A:Authors: Kaine, B.P.; Borodovsky, M.; Kleck, H.P.; Frazer, C.M.; Smith, H.O.; Woese, C.

A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii

A:Reference number: A64300; MID:96337999; PMID:8688087

A:Accession: B64302

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-524 <BUL>

A:Cross-references: GB:U67460; GB:L77117; NID:91590820; PIDN:AA09002.1; PID:91498776; C:Genetics:

A:Map position: FOR17658-19232

Query Match 46.9%; Score 46; DB 2; Length 524;

Best local Similarity 41.2%; Pred. No. 17;

Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

OY 3 LSGIASGRLYISRMNG 19

DB 384 IGNIAGGKLYDNVIG 400

RESULT 4

A39339

protein C inhibitor precursor - human

N:Alternate names: GP43; plasminogen activator inhibitor 3 precursor (PAI3); protein DKF

C:Species: Homo sapiens (man)

C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 20-Jun-2000

C:Accession: A39339; A37341; A26139; G02081; A60434; A60462; S41213; T12502; S41214; S41

J. Biol. Chem. 266, 15028-15034, 1991

A:Title: Characterization of the gene coding for human protein C inhibitor (plasminogen acti

A:Reference number: A39339; MID:91332018; PMID:1714450

A:Accession: A39339

A:Molecule type: DNA

A:Residues: 1-406 <MEI1>

A:Cross-references: GB:M68516; GB:M64881; NID:9189677; PIDN:AAA02811.1; PID:9189678; GB:

R:Meljers, J.C.M.; Chung, D.W.

Thromb. Res. 59, 389-393, 1990

A:Title: Evidence for a glycine residue at position 316 in human protein C inhibitor.

A:Reference number: A37341; MID:91048502; PMID:2173165

A:Accession: A37341

A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra

A:Molecule type: DNA

A:Residues: 331-341 <MEI2>

A:Cross-references: GB:M68516; GB:M64881; NID:9189677

R:Suzuki, K.; Dayashiki, Y.; Nishiohara, J.; Kurachi, K.; Akita, M.; Yamamoto, S.; Hashim

J. Biol. Chem. 262, 611-616, 1987

A:Title: Characterization of a cDNA for human protein C inhibitor.

A:Reference number: A26139; MID:87109153; PMID:3027058

A:Accession: A26139

A:Molecule type: mRNA

A:Residues: 1-334, R, 336-406 <SUZ>

A:Cross-references: GB:J02639; NID:9180549; PIDN:AAA5688.1; PID:9180550

R:Radtke, K.

submitted to the EMBL Data Library, August 1995

A:Reference number: H00782

A:Accession: G02081

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-54, 'V', 56-104, 'E', 106-406 <RAD>

A:Cross-references: EMBL:U5464; NID:91144560; PIDN:AA060386.1; PID:91144561

A:Experimental source: allelic variant B

R:Laurell, M.; Stenflo, J.

Thromb. Haemost. 62, 885-891, 1989

A:Title: Protein C inhibitor from human plasma: characterization of native and cleave

A:Reference number: A60434; MID:90085156; PMID:2556811

A:Accession: A60434

A:Molecule type: protein

A:Residues: 20-39 <LAU>

R:Mizuta, K.; Yasuda, T.; Kishi, K.

Biochem. Genet. 27, 731-743, 1989

A:Title: Biochemical and genetic studies on GP43, a 43-kD glycoprotein detected immun

A:Reference number: A60462; MID:90179688; PMID:2627198

A:Accession: A60462

A:Molecule type: protein

A:Residues: 377-386 <MI2>

A:Note: this protein was found to be polymorphic in a Japanese test population

R:Christenson, A.; Lilja, H.

Eur. J. Biochem. 220, 45-53, 1994

A:Title: Complex formation between protein C inhibitor and prostate-specific antigen

A:Reference number: S41212; MID:94164172; PMID:7509746

A:Accession: S41213

A:Molecule type: protein

A:Residues: 29-38, 'X', 40-44, 'XX', 47-50, 'X', 52, 377, 'X', 379-398, 'X', 400 <CHR>

R:Pouskka, A.; Kleijn, M.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, June 1999

A:Reference number: Z17525

A:Accession: T12502

A:Molecule type: mRNA

A:Residues: 'E', 29-406 <POU>

A:Cross-references: EMBL:AL080185; NID:95262662; PIDN:CA045766.1; PID:95262663

A:Experimental source: adult testis; clone DKFZP434P131

C:Comment: Inhibits protein C (see PIR:KXHU) and plasminogen activator (see PIR:UKHU

C:Genetics:

A:Gene: GDB:PCI

A:Cross-references: GDB:134739; OMIM:601841

A:Map position: 14q32.1-14q32.1

A:Introns: 207/1; 297/2; 346/3

A:Note: DKFZP434P131.1

C:Superfamily: antithrombin III

C:Keywords: glycoprotein; monomer; polymorphism; serine protease inhibitor

F:1-19/Domain: signal sequence #status predicted <SIG>

F:20-406/Product: protein C inhibitor #status experimental <MAT>

F:249-262, 338/Binding site: carbohydrate (asn) (covalent) #status predicted

F:376/Inhibitory site: Arg (plasminogen activator) #status predicted

Query Match 45.9%; Score 45; DB 2; Length 406;

Best local Similarity 45.0%; Pred. No. 19;

Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

OY 1 ADLSGASGRLYISRMNG 20

DB 331 ADLSGISHNSINQVSEVHR 350

RESULT 5

T10633

hypothetical protein T13K14.70 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 18-Aug-2000

C:Accession: T10633

R:Bevan, M.; Pohl, T.; Weizengger, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Lem

submitted to the Protein Sequence Database, June 1999

A:Reference number: Z16991

A:Accession: T10633

A:Molecule type: DNA

A:Residues: 1-957 <BEV>

A:Cross-references: EMBL:AL080282; GSPDB:GN00062; ATSP:T13K14.70

A:Experimental source: cultivar Columbia; BAC clone T13K14

C:Genetics:

A:Gene: ATSP:T13K14.70



C>Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
 C:Accession: AF1198  
 R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
 D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Esht, H.  
 Science 284, 849-852, 2001  
 A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapk, G.; Madueno, E.; Maitournam, A.; Ma  
 Ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
 A:Title: Comparative genomics of *Listeria species*.  
 A:Reference number: AB1077; PMID:21537279; PMID:11679669  
 A:Accession: AF1198  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-453 <GLA>  
 A:Cross-references: GB:NC\_003210; PIDN:GAC99068.1; PID:g16410392; GSPDB:GN00177  
 A:Experimental source: strain EGD-e  
 C:Genetics:  
 A:Gene: lmo0990

Query Match 44.9%; Score 44; DB 2; Length 453;  
 Best Local Similarity 66.7%; Pred. No. 31;  
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 9 GGRUYISRMKG 20  
 |||||  
 Db 75 GGGYISRLGK 86

## RESULT 11

T10634  
 hypothetical protein T13K14.80 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 18-Aug-2000  
 C:Accession: T10634  
 R:Bevan, M.; Pohl, T.; Weizengger, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Lemcke  
 submitted to the protein Sequence Database, June 1999  
 A:Reference number: Z16991  
 A:Accession: T10634  
 A:Molecule type: DNA  
 A:Residues: 1-870 <BEV>  
 A:Cross-references: EMBL:AL080282; GSPDB:GN00062; ATSP:T13K14.80  
 A:Experimental source: cultivar Columbia; BAC clone T13K14  
 C:Genetics:  
 A:Gene: ATSP:T13K14.80  
 A:Map position: 4  
 A:Introns: 73/3; 105/3; 257/2; 328/2; 635/1; 685/3; 724/3  
 C:Superfamily: Arabidopsis thaliana hypothetical protein T13K14.70

Query Match 44.9%; Score 44; DB 2; Length 870;  
 Best Local Similarity 45.0%; Pred. No. 60;  
 Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 1 ADLSGIASGRLYSRMKG 20  
 |||||  
 Db 242 ADKGLKDSQYMTTRTK 261

## RESULT 12

T02712  
 similar to late embryogenesis abundant proteins [imported] - Arabidopsis thaliana

N:Alternate names: hypothetical protein T18E12.19  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 16-Feb-2001  
 C:Accession: T02712; G84444  
 R:Rounsley, S.D.; Lin, X.; Kaul, S.; Shea, T.P.; Fujii, C.Y.; Mason, T.M.; Shen, M.; Ron  
 submitted to the EMBL Data Library, September 1998  
 A:Description: Arabidopsis thaliana chromosome II BAC T18E12 genomic sequence.  
 A:Reference number: Z14702  
 A:Accession: T02712  
 A>Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1805 <ROU>  
 A:Cross-references: EMBL:AC005313; NID:g3548797; PID:g3548816

A:Experimental source: cultivar Columbia  
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.  
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,  
 euss, D.; Nieman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter  
 Nature 402, 761-768, 1999  
 A:Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.  
 A:Reference number: AB4420; PMID:20083487; PMID:10617197  
 A:Accession: G84444  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1805 <STO>  
 A:Cross-references: GB:AE002093; NID:g3548816; PIDN:AAC34488.1; GSPDB:GN00139  
 C:Genetics:  
 A:Gene: T18E12.19; At2g03140  
 A:Map position: 2  
 A:Introns: 249/2; 341/3; 403/3; 451/1; 611/3; 1414/3; 1471/2; 1523/2; 1603/3; 1707/2

Query Match 44.4%; Score 43.5; DB 2; Length 1805;  
 Best Local Similarity 57.9%; Pred. No. 1.5e+02;  
 Matches 11; Conservative 3; Mismatches 4; Indels 1; Gaps 1;

OY 3 LSGIASGRLYSRMKG 20  
 |||||  
 Db 70 LAGIASGLAVYLSRFFGK 88

## RESULT 13

J00740  
 DNA-directed DNA polymerase (EC 2.7.7.7) III beta chain - *Micrococcus luteus* (fragmen

C:Species: *Micrococcus luteus*, *Micrococcus lysodeikticus*  
 C>Date: 10-Mar-1994 #sequence\_revision 10-Mar-1994 #text\_change 18-Jun-1999  
 C:Accession: J00740  
 R:Fujita, M.Q.; Yoshikawa, H.; Ogasawara, N.  
 Gene 93, 73-78, 1990  
 A:Title: Structure of the dna region of *Micrococcus luteus*: conservation and variati  
 A:Reference number: J00737; PMID:91033019; PMID:2172090  
 A:Accession: J00740  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-310 <FUJ>  
 A:Cross-references: GB:M34006; NID:g149845; PIDN:AAA25316.1; PID:g149849  
 C:Genetics:  
 A:Gene: dnaN  
 A:Superfamily: DNA-directed DNA polymerase III beta chain  
 C:Keywords: DNA replication initiation; nucleotidyltransferase

Query Match 43.9%; Score 43; DB 2; Length 310;  
 Best Local Similarity 81.8%; Pred. No. 30;  
 Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ADLSGIASGCR 11  
 |||||  
 Db 220 ADLYGFASGCR 230

## RESULT 14

G84398  
 phosphate transporter permease [imported] - *Halobacterium* sp. NRC-1

C:Species: *Halobacterium* sp. NRC-1  
 C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C:Accession: G84398  
 R:Ng, W.V.; Kennedy, S.P.; Mahatras, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Laaky  
 ; Leitauer, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja  
 Jung, K.H.; Alam, M.; Freilts, T.  
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
 A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Edhardt, H.; Lowe, T.M.;  
 A:Title: Genome sequence of *Halobacterium* species NRC-1.  
 A:Reference number: AB4160; PMID:20504483; PMID:11016950  
 A:Accession: G84398  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-371 <STO>  
 A:Cross-references: GB:AE004437; NID:g10581881; PIDN:AG20555.1; GSPDB:GN00138





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:43:41 ; Search time 1.91268 Seconds  
(without alignments)  
433.698 Million cell updates/sec

Title: US-10-091-442-6

Perfect score: 98

Sequence: 1 ADLSTGASGRLYSRMNGK 20

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues

112892

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt\_40.\*

Listing first 45 summaries

# SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	90	91.8	380	1 SPB7_HUMAN	O75635 homo sapien
2	53	54.1	405	1 SPI2_MOUSE	O91x88 mus musculu
3	51	52.0	378	1 PT16_BOVIN	O02739 bos taurus
4	47	48.0	405	1 SB12_HUMAN	O96683 homo sapien
5	46	46.9	524	1 Y018_MERJA	O60324 methanococc
6	45	45.9	406	1 IPSP_HUMAN	P05134 homo sapien
7	44.5	45.4	433	1 COB3_PSEDE	P21632 pseudomonas
8	43	43.9	310	1 DP3B_MICLU	P21174 micrococcus
9	43	43.9	379	1 ILEU_HORSE	P05619 equus caball
10	43	43.9	391	1 SB13_HUMAN	O91x88 homo sapien
11	43	43.9	405	1 SPI2_HUMAN	O75830 homo sapien
12	43	43.9	622	1 YF54_MERJA	O38949 methanococc
13	42	42.9	126	1 RBF4_TREPA	O83860 treponema p
14	42	42.9	379	1 ILEU_HUMAN	P30740 homo sapien
15	42	42.9	396	1 SX11_CHICK	P48435 gallus gall
16	42	42.9	417	1 KBP_MOUSE	P29621 mus musculu
17	42	42.9	739	1 CAT4_MYCSM	O59557 mycobacteri
18	41	41.8	378	1 ILEU_PIG	P80229 sus scrofa
19	41	41.8	382	1 OVAL_COTJA	P19104 coturnix co
20	41	41.8	386	1 VAB2_GOSHI	O33432 gossypium h
21	41	41.8	392	1 SB11_HUMAN	O96683 homo sapien
22	41	41.8	397	1 SB10_HUMAN	P48595 homo sapien
23	41	41.8	413	1 HP55_TAMSI	O09055 tamias sibi
24	41	41.8	418	1 COTR_MOUSE	P07759 mus musculu
25	41	41.8	423	1 AACT_HUMAN	P01011 homo sapien
26	41	41.8	433	1 VAB2_HORVU	O40079 hordeum vul
27	41	41.8	488	1 VAB1_GOSHI	O33432 gossypium h
28	41	41.8	488	1 VAB1_HORVU	O40078 hordeum vul
29	41	41.8	492	1 VAB2_HORVU	P11574 arabidopsis
30	41	41.8	513	1 VAB1_BOVIN	P14107 bos taurus
31	41	41.8	513	1 VAB1_HUMAN	P15313 homo sapien
32	41	41.8	1322	1 SAL1_MOUSE	O9er14 mus musculu
33	40	40.8	283	1 PSTA_RHILU	O98fi4 rhizobium l

34	40	40.8	328	1 GRHR_HORSE	O18821 equus caball
35	40	40.8	328	1 GRHR_PIG	P49922 sus scrofa
36	40	40.8	378	1 PT16_MOUSE	O60854 mus musculu
37	40	40.8	420	1 DCDA_AOUAE	O67262 aquifex aeo
38	40	40.8	453	1 VAB2_CHICK	P49712 gallus gall
39	40	40.8	485	1 VAB2_TRYCO	O26976 trypanosoma
40	40	40.8	490	1 VAB2_DROME	P31409 drosophila
41	40	40.8	491	1 VAB2_CAEEL	O19626 caenorhabdit
42	40	40.8	492	1 OPU2_BACSU	O06493 bacillus su
43	40	40.8	492	1 VAB1_ACEAT	O38681 acetabulari
44	40	40.8	492	1 VAB2_ACEAT	O38680 acetabulari
45	40	40.8	494	1 VAB2_HELVI	P31410 heliothis v

## ALIGNMENTS

RESULT 1  
ID SPB7\_HUMAN STANDARD: PRT: 380 AA.  
AC O75635;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DE Megsin (TP55) (Serpin B7).  
GN SERPINB7.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97326116; PubMed=9182567;  
RA Tsujimoto M., Tsunooka N., Ishida N., Kurihara T., Iwasa F.,  
Yamashiro K., Rogi T., Kodama S., Katsuragi N., Adachi M.,  
Katayama T., Nakao M., Yamachi K., Hashino J., Haryama M., Miura K.,  
Nakanishi T., Nakazato H., Teramura M., Mizoguchi H., Yamaguchi N.;  
RT "Purification, cDNA cloning, and characterization of a new serpin with  
megakaryocyte maturation activity.";  
RL J. Biol. Chem. 272:15373-15380(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98376492; PubMed=9710452;  
RA Miyata T., Nangaku M., Suzuki D., Inagi R., Uragami K., Sakai H.,  
Okubo K., Kurokawa K.;  
RT "A mesangium-predominant gene, megasin, is a new serpin upregulated in  
IgA nephropathy.";  
RL J. Clin. Invest. 102:828-836(1998).  
CC -!- FUNCTION: Might function as an inhibitor of Lys-specific  
proteases. Might influence the maturation of megakaryocytes via  
its action as a serpin.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -!- TISSUE SPECIFICITY: Predominantly expressed in mesangial cells.  
CC -!- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
CC EMBL: D88575; BAA31232.1; -  
CC EMBL: AF027866; AAC64506.1; -  
CC HSSP: P05619; IHLF.  
CC Genew: HGNC:13902; SERPINB7.  
CC MIM: 603357; -  
CC InterPro: IPR000215; Serpin.  
CC Pfam: PF00079; serpin.1.  
CC SMART: SM00093; SERPIN.1.  
CC PROSITE: PS00284; SERPIN.1.

KW Serpin; Serine protease inhibitor.  
 FT ACT\_SITE 347 348 REACTIVE\_BOND (BY SIMILARITY).  
 SQ SEQUENCE 380 AA: 42904 MW: 9A2CDB6C63CF605 CRC64;  
 Query Match 91.8%; Score 90; DB 1; Length 380;  
 Best Local Similarity 95.0%; Pred. No. 2.5e-07;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ADLGIASGRLYISRMK 20  
 DB 305 ADLGIASGRLYISRMK 324

RESULT 2  
 ID SPI2\_MOUSE STANDARD; PRT; 405 AA.  
 AC O9JX88; Q9D955; Q9D8Z3;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Serpin I2 precursor (Serine protease inhibitor 14).  
 GN SERPINI2 OR SPI14.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Pancreas;  
 RA Chang W.S., Lin S.C., Wu C.W.;  
 RT "Isolation and characterization of mouse pancreas-specific serpin  
 gene.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Pancreas;  
 RX MEDLINE=2108560; PubMed=11217851;  
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamada I.,  
 RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nakai I., Pesole G., Quackenbush J.,  
 RA Schmitt L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Mashio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsch G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hornum M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Noridone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilmung L.,  
 RA Wyszynski A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 CC -1- SUBCELLULAR LOCATION: Secreted (Probable).  
 CC -1- TISSUE SPECIFICITY: Expressed in pancreas.  
 CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

DR EMBL: AF251276; AAF65821.1; -;  
 DR EMBL: AK007347; BAB24976.1; -;  
 DR EMBL: AK007510; BAB25079.1; -;  
 DR HSSP: P01008; IATN.

DR MGD; MGI:894672; SPI14.  
 DR InterPro: IPR000215; Serpin.  
 DR Pfam: PF00079; serpin.1.  
 DR SMART: SM00093; SERPIN.1.  
 DR PROSITE: PS00284; SERPIN.1.  
 KW Serpin; Serine protease inhibitor; Glycoprotein; Signal.  
 FT SIGNAL 1 18  
 FT CHAIN 19 405  
 FT ACT\_SITE 357 358 REACTIVE\_BOND (BY SIMILARITY).  
 FT CARBOHYD 306 306 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 5 5 I -> M (IN REF. 2; BAB25079).  
 FT CONFLICT 195 195 K -> T (IN REF. 2; BAB24976).  
 FT CONFLICT 207 207 D -> S (IN REF. 2; BAB24976).  
 FT CONFLICT 212 218 KYPMMKA -> RVPTEDEV (IN REF. 2; BAB24976).  
 FT CONFLICT 246 246 F -> Y (IN REF. 2; BAB24976).  
 SQ SEQUENCE 405 AA: 45775 MW: 04DF38BE8545DEA8 CRC64;

Query Match 54.1%; Score 53; DB 1; Length 405;  
 Best Local Similarity 52.6%; Pred. No. 0.31;  
 Matches 10; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 2 DLGIASGRLYISRMK 20  
 DB 316 DLGIASGRLYISRMK 334

RESULT 3  
 ID PT16\_BOVIN STANDARD; PRT; 378 AA.  
 AC O02739;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Serine proteinase inhibitor B-43.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxId=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=97165893; PubMed=9013786;  
 RA Nakaya N., Nishibori M., Kawabata M., Saeki K.;  
 RT "Cloning of a serine proteinase inhibitor from bovine brain:  
 RT expression in the brain and characterization of its target  
 RT proteinases.";  
 RL Brain Res. Mol. Brain Res. 42:293-300(1996).  
 CC -1- FUNCTION: MAY BE INVOLVED IN THE REGULATION OF SERINE PROTEINASES  
 CC PRESENT IN THE BRAIN OR EXTRAVALSATED FROM THE BLOOD.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- TISSUE SPECIFICITY: BRAIN.  
 CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

DR EMBL: D55670; BAA19875.1; -;  
 DR HSSP: P05120; IBY7.  
 DR InterPro: IPR000215; Serpin.  
 DR Pfam: PF00079; serpin.1.  
 DR SMART: SM00093; SERPIN.1.  
 DR PROSITE: PS00284; SERPIN.1.  
 KW Serpin; Serine protease inhibitor.  
 FT ACT\_SITE 343 344 REACTIVE\_BOND (BY SIMILARITY).  
 SQ SEQUENCE 378 AA: 42560 MW: 66AF49CCCE263A CRC64;

Query Match 52.0%; Score 51; DB 1; Length 378;

Best Local Similarity 50.0%; Pred. No. 0.62;  
Matches 10; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

OY 1 ADLSGASGRLYISRMKG 20  
DB 301 ADLSGASGRLYISRMKG 320

## RESULT 4

ID SB12\_HUMAN STANDARD; PRT; 405 AA.  
AC 096P63;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Serpin B12.  
GN SERPINB12.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN 11  
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
RX MEDLINE=21623618; PubMed=11604408;  
RA Askew Y.S., Pak S.C., Luke C.J., Askew D.J., Cataltepe S., Mills D.R.,  
RA Kato H., Lehoccky J., Dewar K., Birren B., Silverman G.A.;  
RT \*SERPINB12 is a novel member of the human ov-serpin family that is  
RT widely expressed and inhibits trypsin-like serine proteinases.";  
RL J. Biol. Chem. 276:49320-49330(2001).  
CC 1- FUNCTION: Inhibits trypsin and plasmin, but not thrombin,  
CC coagulation factor Xa, or urokinase-type plasminogen activator.  
CC 1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC 1- TISSUE SPECIFICITY: Expressed in many tissues, including brain,  
CC bone marrow, lymph node, heart, lung, liver, pancreas, testis,  
CC ovary, and intestine.  
CC 1- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: AF411191; AAL05571.1; .  
DR Genew: HGNC:14220; SERPINB12.  
DR InterPro: IPR000215; Serpin.  
DR Pfam: PF00079; serpin.1.  
DR PROSITE: PS00284; SERPIN.1.  
KW Serpin, Serine protease inhibitor.  
FT ACT\_SITE 370 371 REACTIVE\_BOND (BY SIMILARITY).  
SQ SEQUENCE 405 AA; 46276 MW; FFE12D4C9B7F3DFA CRC64;

Query Match 48.0%; Score 47; DB 1; Length 405;  
Best Local Similarity 45.0%; Pred. No. 3;  
Matches 9; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

OY 1 ADLSGASGRLYISRMKG 20  
DB 328 ADLSGASGRLYISRMKG 347

## RESULT 5

ID Y018\_METJA STANDARD; PRT; 524 AA.  
AC 060324;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein M70018.  
GN M70018.  
OS Methanococcus jannaschii.

OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
OC Methanocaldococcaceae; Methanocaldococcus.  
OX NCBI\_TaxID=2190;  
RN 11  
RP SEQUENCE FROM N.A.  
RX STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
RX MEDLINE=96337999; PubMed=8688087;  
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reisch C.I.,  
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,  
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
RA Cotton M.D., Roberts K.M., Hurd M.A., Kaine B.P., Bordovsky M.,  
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
RT jannaschii.";  
RL Science 273:1058-1073(1996).

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL: U67460; AAB98002.1; .  
DR TIGR: M70018; .  
KW Hypothetical protein; Transmembrane; Complete proteome.  
FT TRANSMEM 13 33 POTENTIAL.  
SQ SEQUENCE 524 AA; 56565 MW; 0F878499773FF87C CRC64;

Query Match 46.9%; Score 46; DB 1; Length 524;  
Best Local Similarity 41.2%; Pred. No. 5.7;  
Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

OY 3 LSGIAGSGRLYISRMKG 19  
DB 384 LSGIAGSGRLYISRMKG 400

## RESULT 6

ID IPSP\_HUMAN STANDARD; PRT; 406 AA.  
AC P05154; Q9UG30; Q07616;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Plasma serine protease inhibitor precursor (PCI) (Protein C inhibitor)  
DE (Plasminogen activator inhibitor-3) (PAI3) (Acrosomal serine protease  
DE inhibitor).  
GN SERPINAS OR PCI OR PLANH3 OR PROCI.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN 11  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87109153; PubMed=3027058;  
RA Suzuki K., Devashiki Y., Nishiohara J., Kurachi K., Akira M.,  
RA Yamamoto S., Hashimoto S.;  
RT "Characterization of a cDNA for human protein C inhibitor. A new  
RT member of the plasma serine protease inhibitor superfamily.";  
RL J. Biol. Chem. 262:611-616(1987).  
RN 12  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91048502; PubMed=2173165;  
RA Meijers J.C.M., Chung D.W.;  
RT "Evidence for a glycine residue at position 316 in human protein C  
RT inhibitor.";  
RL Thromb. Res. 59:389-393(1990).  
RN 13

RP SEQUENCE FROM N.A.  
 RX MEDLINE-91332018; PubMed-1714450;  
 RA Meljers J.C.M., Chung D.W.;  
 RT "Organization of the gene coding for human protein C inhibitor  
 (plasminogen activator inhibitor-3). Assignment of the gene to  
 chromosome 14.";  
 RL J. Biol. Chem. 266:15028-15034(1991).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-94198434; PubMed-8148499;  
 RA Hayashi T., Suzuki K.;  
 RT "Gene organization of human protein C inhibitor, a member of SERPIN  
 family proteins encoded in five exons.";  
 RL Int. J. Hematol. 58:213-224(1993).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Liver;  
 RA Radtke K.P., Greengard J.S., Fernandez J.A., Villoutreix B.O.,  
 RL Griffin J.H.;  
 RT Submitted (Jan-1996) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A., AND VARIANTS G-44; V-55; S-64; V-94; E-105; P-115  
 AND R-217.  
 RA Rieder M.J., Carrington D.P., Chung M.-W., Lee K.L., Poel C.L., Yi Q.,  
 RL Nickerson D.A.;  
 RT Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Skin;  
 RA Strausberg R.;  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 RN [8]  
 RP SEQUENCE OF 27-406 FROM N.A.  
 RC TISSUE-Testis;  
 RX MEDLINE-93329000; PubMed-8471250;  
 RA Moore A., Penfold L.M., Johnson J.L., Latchman D.S., Moore H.D.;  
 RT "Human sperm-egg binding is inhibited by peptides corresponding to  
 core region of an acrosomal serine protease inhibitor.";  
 RL Mol. Reprod. Dev. 34:280-291(1993).  
 RN [9]  
 RP SEQUENCE OF 28-406 FROM N.A.  
 RC TISSUE-Testis;  
 RA Poustka A., Klein M., Mewes H.-W., Gassenhuber J., Wiemann S.;  
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
 RN [10]  
 RP SEQUENCE OF 20-39.  
 RX MEDLINE-90085156; PubMed-2556811;  
 RA Laurell M., Stenflo J.;  
 RT "Protein C inhibitor from human plasma: characterization of native  
 and cleaved inhibitor and demonstration of inhibitor complexes with  
 plasma kallikrein.";  
 RL Thromb. Haemost. 62:885-891(1989).  
 RN [11]  
 RP 3D-STRUCTURE MODELING.  
 RX MEDLINE-91046026; PubMed-2172989;  
 RA Kuhn L.A., Griffin J.H., Fisher C.L., Greengard J.S., Bouma B.N.,  
 RA Espana F., Tainer J.A.;  
 RT "Elucidating the structural chemistry of glycosaminoglycan  
 recognition by protein C inhibitor.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:8506-8510(1990).  
 CC -I- FUNCTION: INHIBITS ACTIVATED PROTEIN C AS WELL AS PLASMINOGEN  
 CC ACTIVATORS.  
 CC -I- SUBCELLULAR LOCATION: Secreted.  
 CC -I- TISSUE SPECIFICITY: SYNTHESIZED IN LIVER; SECRETED IN PLASMA.  
 CC -I- DISEASE: A DEFICIENCY OF PROTEIN C INHIBITOR IS THE CAUSE OF A  
 CC COMBINED FACTOR V/VIII DEFICIENCY DISEASE.  
 CC -I- SIMILARITY: BELONGS TO THE SERPIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M68516; AAA02811.1; -  
 DR EMBL: J02639; AAA35688.1; -  
 DR EMBL: S69366; AAB30461.1; -  
 DR EMBL: S69364; AAB30461.1; JOINED.  
 DR EMBL: S69364; AAB30461.1; JOINED.  
 DR EMBL: S69365; AAB30461.1; JOINED.  
 DR EMBL: AF361796; AAK27240.1; -  
 DR EMBL: U35464; AAB60386.1; -  
 DR EMBL: BC008915; AAH08915.1; -  
 DR EMBL: S58545; AAB26244.2; -  
 DR EMBL: AL080185; CAB45766.1; -  
 DR PIR: A39339; A39339.  
 DR PDB: 1PAI; 15-OCT-94.  
 DR PDB: 2PAI; 15-OCT-94.  
 DR Genew; HGNC:8723; SERPIN5.  
 DR MIM: 601841; -  
 DR InterPro: IPR000215; Serpin.  
 DR Pfam: PF00079; serpin; 3.  
 DR SMART: SM00093; SERPIN; 1.  
 DR PROSITE: PS00284; SERPIN; 1.  
 KW Serpin; Serine protease inhibitor; Plasma; Glycoprotein; Signal;  
 KW Polymorphism; 3D-structure.  
 FT SIGNAL 1 19  
 FT CHAIN 20 406  
 FT CARBOHYD 249 249  
 FT CARBOHYD 262 262  
 FT CARBOHYD 338 338  
 FT ACT\_SITE 373 374  
 FT VARIANT 44 44  
 FT FT  
 FT VARIANT 55 55  
 FT FT  
 FT VARIANT 64 64  
 FT FT  
 FT VARIANT 94 94  
 FT FT  
 FT VARIANT 105 105  
 FT FT  
 FT VARIANT 115 115  
 FT FT  
 FT VARIANT 217 217  
 FT FT  
 FT CONFLICT 28 28  
 FT CONFLICT 221 221  
 FT CONFLICT 335 335  
 FT CONFLICT 384 384  
 SQ SEQUENCE 406 AA; 45701 MW; 0B9D1A519341B8C9 CRC64;  
 Query Match 45.9%; Score 45; DB 1; Length 406;  
 Best Local Similarity 45.0%; Pred. No. 6.4;  
 Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;  
 QY 1 ADLGSIASGRLYSRMGK 20  
 DB 331 ADLGSINSHNINQSEVHWK 350  
 |||||: : : : :  
 RESULT 7  
 COBB\_PSEDE  
 ID COBB\_PSEDE STANDARD; PRT; 433 AA.  
 AC P21632;  
 DT 01-MAY-1991 (Rel. 18, Created)  
 DT 01-MAY-1991 (Rel. 18, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Cobyrinic acid A,C-diamide synthase.  
 GN COBB.  
 OS Pseudomonas denitrificans.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=43306;

RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-15.  
 RC STRAIN-SC510;  
 RX MEDLINE-91008975; PubMed-2211520;  
 RA Rouzet J., Cauchat L., Blanche F., Debussche L., Thibaut D.,  
 RY Rouzet J., Cauchat L., Blanche F., Debussche L., Thibaut D.,  
 RT "Nucleotide sequence of a pseudomonas denitrificans 5.4-kilobase DNA  
 fragment containing five cob genes and identification of structural  
 RT genes encoding S-adenosyl-L-methionine: uroporphyrinogen III  
 methyltransferase and cobyrinic acid a,c-dimide synthase.";  
 RL J. Bacteriol. 172:5968-5979(1990).  
 CC -1- FUNCTION: RESPONSIBLE FOR THE AMIDATION OF CARBOXYLIC GROUPS AT  
 CC POSITION A AND C OF EITHER COBIRINIC ACID OR HYDROXYBIRINIC ACID.  
 CC NH2(2) GROUPS ARE PROVIDED BY GLUTAMINE, AND ONE MOLECULE OF ATP  
 CC IS HYDROGENOLYZED FOR EACH AMIDATION.  
 CC -1- PATHWAY: cobalamin biosynthesis.  
 CC -1- SUBUNIT: cobalamin biosynthesis.  
 CC -1- SIMILARITY: BELONGS TO THE COBB/COBQ FAMILY. COBB SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: M59236; AAA25774.1; -  
 DR PIR: C36144; C36144.  
 DR InterPro: IPR004484; CbIA.  
 DR InterPro: IPR002586; CbIA\_P.  
 DR Pfam: PF01656; CbIA; 1.  
 DR TIGRFAMS: TIGR00379; cbia; 1.  
 KM Cobalamin biosynthesis; Porphyryn biosynthesis.  
 FT INIT\_MER  
 FT SEQUENCE 433 AA; 45574 MW; CCD2F56B9698E6DA CRC64;  
 SQ  
 Query Match 45.4%; Score 44.5; DB 1; Length 433;  
 Best Local Similarity 55.0%; Pred. No. 8.3;  
 Matches 11; Conservative 3; Mismatches 5; Indels 1; Gaps 1;  
 QY 1 ADLGSIGASGR-LYISRMWG 19  
 Db 70 ANASHVAGSGRTLIVEAMWG 89  
 ID DPJB\_MICLU STANDARD: PRT; 310 AA.  
 AC P21174;  
 DT 01-MAY-1991 (Rel. 18, Created)  
 DT 01-MAY-1991 (Rel. 18, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE DNA polymerase III, beta chain (EC 2.7.7.7) (Fragment).  
 GN DNAN.  
 OS Micrococcus luteus (Micrococcus lysodeikticus).  
 OC Bacteria; Actinobacteria; Actinobacteriales; Actinobacteridae;  
 OC Actinomycetales; Micrococcales; Micrococcaceae; Micrococcus.  
 OX NCBI\_TaxID=1270;  
 RN RN  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-91033019; PubMed-2172090;  
 RA Fujita M.O., Yoshikawa H., Ogasawara N.;  
 RT "Structure of the dna region of Micrococcus luteus: conservation and  
 RT variations among eubacteria.";  
 RL Gene 93:73-78(1990).  
 CC -1- FUNCTION: DNA POLYMERASE III IS A COMPLEX, MULTICHAIN ENZYME  
 CC RESPONSIBLE FOR MOST OF THE REPLICATIVE SYNTHESIS IN BACTERIA.  
 CC THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.  
 CC THE BETA CHAIN IS REQUIRED FOR INITIATION OF REPLICATION ONCE IT  
 CC IS CLAMPED ONTO DNA, IT SLIDES FREELY (BIDIRECTIONAL AND ATP-  
 CC INDEPENDENT) ALONG DUPLEX DNA (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate -> N diphosphate

CC + (DNA)(N).  
 CC -1- SUBUNIT: CONTAINS A CORE (COMPOSED OF ALPHA, EPSILON, AND  
 CC THERA CHAINS) THAT CAN REPAIR SHORT GAPS CREATED BY NUCLEASE  
 CC IN DUPLEX DNA. FOR EFFICIENT REPLICATION OF THE LONG, SINGLE-  
 CC STRANDED TEMPLATES, POL III REQUIRES THE AUXILIARY CHAINS BETA,  
 CC GAMMA, AND DELTA (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: M34006; AAA25316.1; -  
 DR PIR: JQ0740; JQ0740.  
 DR InterPro: IPR001001; DNA\_polIII\_beta.  
 DR InterPro: IPR001230; Ptenyl\_site.  
 DR Pfam: PF00712; DNA\_pol3\_beta; 1.  
 DR Pfam: PF02767; DNA\_pol3\_beta\_2; 1.  
 DR Pfam: PF02768; DNA\_pol3\_beta\_3; 1.  
 DR SMART: SM00480; POL3bc; 1.  
 DR TIGRFAMS: TIGR00663; dnan; 1.  
 KM Transferase, DNA-directed DNA polymerase; DNA replication.  
 FT NON\_TER  
 FT SEQUENCE 310 AA; 32624 MW; E1BE651CB102F1 CRC64;  
 SQ  
 Query Match 43.9%; Score 43; DB 1; Length 310;  
 Best Local Similarity 81.8%; Pred. No. 10;  
 Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 ADLGSIGASGR 11  
 Db 220 ADLVGFASGR 230  
 ID ILEU\_HORSE STANDARD: PRT; 379 AA.  
 AC P05619;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-NOV-1988 (Rel. 09, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Leukocyte elastase inhibitor (LEI).  
 GN SERPINB1 OR ELANH2.  
 OS Equus caballus (Horse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.  
 OX NCBI\_TaxID=9796;  
 RN RN  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-93319507; PubMed-7687128;  
 RA Kordula T., Dublin A., Schooltink H., Koj A., Heinrich P.C.,  
 RT Rose-John S.;  
 RT "Molecular cloning and expression of an intracellular serpin: an  
 RT elastase inhibitor from horse leukocytes.";  
 RL Biochem. J. 293:187-193(1993).  
 RN RN  
 RP SEQUENCE.  
 RC TISSUE=leukocyte;  
 RX MEDLINE-92202200; PubMed-1551869;  
 RA Dublin A., Travis J., Enghild J.J., Potempa J.;  
 RT "Equine leukocyte elastase inhibitor. Primary structure and  
 RT identification as a thymosin-binding protein.";  
 RL J. Biol. Chem. 267:6576-6583(1992).  
 RN RN  
 RP SEQUENCE OF 343-362.  
 RX MEDLINE-88213423; PubMed-3366785;  
 RA Potempa J., Dublin A., Watorek W., Travis J.;  
 RT "An elastase inhibitor from equine leukocyte cytosol belongs to the  
 RT serpin superfamily. Further characterization and amino acid sequence

RT of the reactive center." ;  
 J. Biol. Chem. 263:7364-7369(1988).  
 [4]  
 RN X-RAY CRYSTALLOGRAPHY (1.95 ANGSTROMS).  
 RP MEDLINE=92389327; PubMed=1518052;  
 RA Baumann U., Bode W., Huber R., Travis J., Potempa J.;  
 RT "Crystal structure of cleaved equine leucocyte elastase inhibitor  
 RT determined at 1.95-A resolution." ;  
 RL J. Mol. Biol. 226:1207-1218(1992).  
 CC -!- FUNCTION: THIS INHIBITOR IS THOUGHT TO BE INVOLVED IN THE CONTROL  
 CC OF INTRACELLULAR PROTEIN TURNOVER. BINDS TO THYMOSIN BETA-4.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M91161; AAA97513.1; -  
 DR PIR: A28060; A28060.  
 DR PIR: A37276; A37276.  
 DR PIR: A42421; A42421.  
 DR PIR: S34062; S34062.  
 DR PDB: 1HE; 3I-JAN-94.  
 DR InterPro: IPR000215; Serpin.  
 DR Pfam: PF00079; serpin; 1.  
 DR SMART: SM00093; SERPIN; 1.  
 DR PROSITE: PS00284; SERPIN; 1.  
 DR Serpin; Serine protease inhibitor; 3D-structure.  
 KM MOD.RES 1 1  
 FT ACT SITE 344 345  
 FT STRAND 2 22  
 FT HELIX 28 30  
 FT HELIX 32 44  
 FT TURN 45 45  
 FT TURN 48 58  
 FT TURN 59 59  
 FT HELIX 60 62  
 FT TURN 64 65  
 FT HELIX 66 67  
 FT TURN 78 78  
 FT STRAND 85 95  
 FT TURN 96 97  
 FT HELIX 102 112  
 FT STRAND 115 119  
 FT TURN 121 123  
 FT HELIX 125 139  
 FT TURN 140 142  
 FT TURN 150 151  
 FT TURN 155 156  
 FT STRAND 157 171  
 FT HELIX 177 179  
 FT STRAND 181 186  
 FT STRAND 192 209  
 FT HELIX 210 212  
 FT TURN 213 213  
 FT STRAND 214 221  
 FT TURN 222 223  
 FT STRAND 226 233  
 FT HELIX 244 248  
 FT TURN 249 249  
 FT HELIX 252 259  
 FT HELIX 261 263  
 FT STRAND 265 274  
 FT STRAND 276 283  
 FT HELIX 285 291  
 FT TURN 292 292  
 FT HELIX 295 297  
 FT TURN 299 301

FT HELIX 305 308  
 FT STRAND 313 326  
 FT STRAND 330 343  
 FT STRAND 350 353  
 FT STRAND 358 364  
 FT TURN 365 368  
 FT STRAND 369 376  
 SQ SEQUENCE 379 AA; 43006 MW; 72DED1999371427A CRC64;  
 Query Match 43 9%; Score 43; DB 1; Length 379;  
 Best Local Similarity 40.0%; Pred. No. 13;  
 Matches 8; Conservative 7; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 ADLGIASGRLYISRMNGK 20  
 DB 302 ADLSGMSGARDLFVSKTIHK 321  
 RESULT 10  
 ID SB13\_HUMAN STANDARD; PRT; 391 AA.  
 AC Q9UYV8; Q9UYV8; Q9UKG0; Q9HCX1;  
 DT 16-OCT-2001 (Rel. 40; Created)  
 DT 16-OCT-2001 (Rel. 40; Last sequence update)  
 DT 15-JUN-2002 (Rel. 41; Last annotation update)  
 DE Hurpin (Hacat UV-repressible serpin) (Protease inhibitor 13)  
 DE (Headpin) (Serpin B13).  
 GN SERPINB13 OR PIL3  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RP [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC TISSUE=Skin;  
 RX MEDLINE=99443898; PubMed=10512713;  
 RA Abts H.F., Weiss T., Mirmohammadsadeh A., Koehrer K., Michel G.,  
 RA Ruzicka T.;  
 RT "Cloning and characterization of hurpin (Protease inhibitor 13): a new  
 RT skin specific, UV-repressible serine proteinase inhibitor of the  
 RT ovalbumin serpin family." ;  
 RL J. Mol. Biol. 293:29-39(1999).  
 RN [2]  
 RN SEQUENCE FROM N.A. (ISOFORM 1).  
 RP MEDLINE=99458661; PubMed=10527881;  
 RA Spring P., Nakashima T., Frederick M., Henderson Y., Clayman G.;  
 RT "Identification and cDNA cloning of headpin, a novel differentially  
 RT expressed serpin that maps to chromosome 18q." ;  
 RL Biochem. Biophys. Res. Commun. 264:299-304(1999).  
 RN [3]  
 RN SEQUENCE FROM N.A. (ISOFORM 1).  
 RP MEDLINE=20461850; PubMed=11004515;  
 RA Nakashima T., Pak S.C., Silverman G.A., Spring P.M., Frederick M.J.,  
 RA Clayman G.L.;  
 RT "Genomic cloning, mapping, structure and promoter analysis of HEADPIN,  
 RT a serpin which is down-regulated in head and neck cancer cells." ;  
 RL Biochim. Biophys. Acta 1492:441-446(2000).  
 RN [4]  
 RN SEQUENCE FROM N.A. (ISOFORM 2).  
 RC TISSUE=Placenta;  
 RX MEDLINE=21213570; PubMed=11313015;  
 RA Abts H.F., Weiss T., Scheuring S., Scott F.L., Irving J.A., Michel G.,  
 RA Bird P.I., Ruzicka T.;  
 RT "Sequence, organization, chromosomal localization and alternative  
 RT splicing of the human serine protease inhibitor gene hurpin (PIL3),  
 RT which is up-regulated in psoriasis." ;  
 RL DNA Cell Biol. 20:123-131(2001).  
 CC -!- FUNCTION: May play a role in the proliferation or differentiation  
 CC of keratinocytes.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; are  
 CC produced by alternative splicing.  
 CC -!- TISSUE SPECIFICITY: SKIN-SPECIFIC.



-1- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.

-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

DR EMBL: AJ001696; CA004935.2; -  
 DR EMBL: AJ001697; CA004936.2; -  
 DR EMBL: AJ001698; CA004937.1; -  
 DR EMBL: AF169949; AAD55765.1; -  
 DR EMBL: AF216854; AAF72879.1; -  
 DR EMBL: AJ278177; CAC03569.1; -  
 DR HSSP: P05120; 1B77.  
 DR Genew: HGNC:8944; SERPINB13.  
 DR MIM: 604445; -  
 DR InterPro: IPR000240; Maspin.  
 DR InterPro: IPR000215; Serpin.  
 DR Pfam: PF00079; serpin.1.  
 DR PRINTS: PRO0676; MASPIN.  
 DR SMART: SM00093; SERPIN.1.  
 DR PROSITE: PS00284; SERPIN.1.  
 DR Serpin; Serine protease inhibitor; Alternative splicing.  
 DR ACT\_SITE 356 357 REACTIVE\_BOND (BY SIMILARITY).  
 FT VARSPIC 206 257 MISSING (IN ISOREF 2).  
 FT CONFLICT 8 8 S -> N (IN REF. 4).  
 FT CONFLICT 75 75 MISSING (IN REF. 1; CA004937).  
 FT CONFLICT 293 293 G -> S (IN REF. 2).  
 FT CONFLICT 297 297 E -> Q (IN REF. 1; CA004937).  
 SO SEQUENCE 391 AA; 44276 MW; 2CAB8558DBAC2B09 CRC64;

Query Match 43.9%; Score 43; DB 1; Length 391;  
 Best Local Similarity 44.4%; Pred. No. 13;  
 Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 ADLSGIASGRLYSRM 18  
 DB 314 ADLSGMSGSLYAKFL 331

RESULT 11  
 SP12\_HUMAN STANDARD; PRT; 405 AA.  
 AC 075830;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Serpin I2 precursor (Myoepithelium-derived serine protease inhibitor)  
 DE (Pancrep) (Protease Inhibitor 14) (TSA2004).  
 GN SERPIN2 OR P114 OR MEPI.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX PubMed-9624529;  
 RA Oaki K., Nagata M., Suzuki M., Fujiwara T., Miyoshi Y., Ishikawa O.,  
 RA Ohigashi H., Imoka S., Takahashi E., Nakamura Y.,  
 RA "Isolation and characterization of a novel human pancreas-specific  
 RT gene, pancpn, that is down-regulated in pancreatic cancer cells.";  
 RL Genes Chromosomes Cancer 22:179-185(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX PubMed-10097100;  
 RA Xiao G., Liu Y.E., Gentz R., Sang Q.A., Ni J., Goldberg I.D.,  
 RA Shi Y.E.;  
 RT "Suppression of breast cancer growth and metastasis by a serpin  
 RT myoepithelium-derived serine proteinase inhibitor expressed in the  
 RT mammary myoepithelial cells.";

RL Proc. Natl. Acad. Sci. U.S.A. 96:3700-3705(1999).  
 CC -1- SUBCELLULAR LOCATION: Secreted (Probable).  
 CC -1- TISSUE SPECIFICITY: Expressed in pancreas and adipose tissues.  
 CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.

-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

DR EMBL: AB006423; BAA33766.1; -  
 DR EMBL: AF130470; AAD34723.1; -  
 DR HSSP: P05120; 1B77.  
 DR Genew: HGNC:8945; SERPINI2.  
 DR MIM: 605587; -  
 DR InterPro: IPR000215; Serpin.  
 DR Pfam: PF00079; serpin.1.  
 DR SMART: SM00093; SERPIN.1.  
 DR PROSITE: PS00284; SERPIN.1.  
 DR Serpin; Serine protease inhibitor; Glycoprotein; Signal.  
 FT SIGNAL 1 18 POTENTIAL.  
 FT CHAIN 19 405 SERPIN I2.  
 FT ACT\_SITE 357 358 REACTIVE\_BOND (BY SIMILARITY).  
 FT CARBOHYD 202 202 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 207 207 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 306 306 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SO SEQUENCE 405 AA; 46145 MW; 5BA18C60E4FDE9A4 CRC64;

Query Match 43.9%; Score 43; DB 1; Length 405;  
 Best Local Similarity 42.1%; Pred. No. 14;  
 Matches 8; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 2 DLSGIASGRLYSRM 20  
 DB 316 DLSGIDSSSEVYVSQYTK 334

RESULT 12  
 YF54\_METJA STANDARD; PRT; 622 AA.  
 AC 058949;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein MJ1554 precursor.  
 GN MJ1554.  
 OS Methanococcus jannaschii.  
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
 OC Methanocaldococcaceae; Methanocaldococcus.  
 OX NCBI\_TaxID=2190;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SRRAIN-JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE-96337999; PubMed-8688087;  
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,  
 RA Utermack T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Bordovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 RT jannaschii.";  
 RL Science 273:1058-1073(1996).  
 CC -1- SIMILARITY: TO A.FULGIDUS AF0817.

-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its



DR SMART; SM00093; SERPIN; 1.  
 DR PROSITE; PS00284; SERPIN; 1.  
 KM Serpin; Serine protease inhibitor.  
 FT ACT\_SITE 344 345  
 SQ SEQUENCE 379 AA; 42741 MW; BAAE08DFC8D8CD3 CRC64;

## Query Match

Best Local Similarity 42.9%; Score 42; DB 1; Length 379;

Matches 8; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

QY 1 ADLGGIAGGRLYISRMGK 20

DB 302 ADLGGMGARDIFISKIVHK 321

## RESULT 15

SK11\_CHICK STANDARD; PRT; 396 AA.

ID SK11\_CHICK

AC P48435; 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Transcription factor SOX-11.

GN SOX11.

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauaria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.

OX NCBI\_TaxID=9031;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Embryo;

RA MEDLINE=95267693; PubMed=7748786;

RA Uvanogho D., Rex M., Cartwright E.J., Pearl G., Healy C.,

RA Scotling P.J., Sharpe P.T.;

RT "Embryonic expression of the chicken Sox2, Sox3 and Sox11 genes

RT suggests an interactive role in neuronal development.";

RL Mech. Dev. 49:23-36(1995).

CC -1- FUNCTION: MAY FUNCTION AS SWITCHES IN NEURONAL DEVELOPMENT.

CC -1- SUBCELLULAR LOCATION: Nuclear.

CC -1- TISSUE SPECIFICITY: LOW LEVEL EXPRESSION IS SEEN IN

CC UNDIFFERENTIATED PROLIFERATING CELLS OF NEURAL EPITHELIUM. A

CC GREATER EXPRESSION IS SEEN IN THE MATURING NEURONS AFTER THEY

CC LEAVE THE NEURAL EPITHELIUM. IT IS ALSO FOUND IN THE GUT

CC EPITHELIUM AND ADRENAL MEDULLA.

CC -1- SIMILARITY: CONTAINS 1 HMG BOX.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC -----

CC EMBL: U12534; AAB09664.1; -

DR HSSP; P48436; 1SX9

DR InterPro; IPR000910; HMG\_12\_box.

DR Pfam; PF00505; HMG\_box; 1.

DR SMART; SM00398; HMG; 1.

KM Transcription regulation; DNA-binding; Nuclear protein.

FT DNA\_BIND 49 117

FT DOMAIN 162 165

FT DOMAIN 204 212

FT DOMAIN 288 294

FT DOMAIN 332 335

SQ SEQUENCE 396 AA; 43503 MW; 8E4B0A57F8BA833 CRC64;

DB 266 EVRGAGGRLTYIS 279

Search completed: July 11, 2003, 11:55:17  
 Job time : 2.91268 secs

Query Match 42.9%; Score 42; DB 1; Length 396;

Best Local Similarity 57.1%; Pred. No. 19;

Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 DLGGIAGGRLYIS 15

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:44:31; Search time 8.73181 Seconds

(without alignments)  
471.946 Million cell updates/sec

Title: US-10-091-442-6

Perfect score: 98

Sequence: 1 ADLSCIASGRLYSMMGK 20

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :  
1: sp.archaea:\*  
2: sp.bacteria:\*  
3: sp.fungi:\*  
4: sp.human:\*  
5: sp.invertebrate:\*  
6: sp.mammal:\*  
7: sp.mhc:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp.rodent:\*  
12: sp.virus:\*  
13: sp.vertibrate:\*  
14: sp.unclassified:\*  
15: sp.rvivirus:\*  
16: sp.bacteriap:\*  
17: sp.archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	83	84.7	360	11	Q9D695	Q9D695 mus musculus
2	83	84.7	360	11	Q9D20J5	Q9D20J5 rattus norv
3	54	55.1	191	11	Q63547	Q63547 rattus norv
4	50	51.0	615	4	Q9NXX3	Q9NXX3 homo sapien
5	50	51.0	615	4	Q8TAP0	Q8TAP0 homo sapien
6	48	49.0	423	11	Q9D7P9	Q9D7P9 mus musculus
7	48	49.0	591	16	Q9RW57	Q9RW57 delnoccocus
8	46	46.9	327	6	Q9M2I6	Q9M2I6 canis fam11
9	46	46.9	377	11	Q08804	Q08804 mus musculus
10	46	46.9	382	5	Q9GPI2	Q9GPI2 schistosoma
11	46	46.9	400	5	Q967L9	Q967L9 schistosoma
12	45	45.9	373	5	Q9NH65	Q9NH65 trichinella
13	45	45.9	385	5	Q26509	Q26509 schistosoma
14	45	45.9	942	10	Q9C509	Q9C509 arabidopsis
15	45	45.9	942	10	Q9C508	Q9C508 arabidopsis
16	45	45.9	942	10	Q9A5R3	Q9A5R3 arabidopsis

17	45	45.9	957	10	Q9SUC2	Q9SUC2 arabidopsis
18	44	44.9	205	16	Q31560	Q31560 bacillus su
19	44	44.9	222	17	Q9YEB7	Q9YEB7 aeropyrum p
20	44	44.9	377	5	Q9GPI3	Q9GPI3 ixodes ric1
21	44	44.9	379	11	Q9D7S8	Q9D7S8 mus musculu
22	44	44.9	379	11	Q9D154	Q9D154 mus musculu
23	44	44.9	386	13	Q73860	Q73860 meleagris g
24	44	44.9	388	11	Q9COV3	Q9COV3 mus musculu
25	44	44.9	388	11	Q91212	Q91212 mus musculu
26	44	44.9	433	16	Q9ZD31	Q9ZD31 listeria in
27	44	44.9	453	16	Q8Y8B8	Q8Y8B8 listeria mo
28	44	44.9	453	16	Q8Y8B8	Q8Y8B8 listeria mo
29	43.5	44.4	293	2	Q9SUC1	Q9SUC1 arabidopsis
30	43.5	44.4	293	2	Q9LBR0	Q9LBR0 porphyromon
31	43.5	44.4	431	12	Q910Z6	Q910Z6 rana tigrin
32	43.5	44.4	431	12	Q9YKK2	Q9YKK2 epizootic h
33	43.5	44.4	431	12	Q8Q202	Q8Q202 rana tigrin
34	43.5	44.4	1805	10	Q81060	Q81060 arabidopsis
35	43	43.9	203	6	Q9GMA8	Q9GMA8 sus scrofa
36	43	43.9	231	6	Q28500	Q28500 macaca mula
37	43	43.9	309	17	Q97B69	Q97B69 thermoplasma
38	43	43.9	359	17	Q8TJ67	Q8TJ67 methanocarc
39	43	43.9	371	17	Q9HML6	Q9HML6 halobacteri
40	43	43.9	378	10	Q8S2S5	Q8S2S5 thelunglei
41	43	43.9	382	11	Q8VHP7	Q8VHP7 mus musculu
42	43	43.9	416	16	Q98KS9	Q98KS9 rhizobium l
43	43	43.9	418	11	Q62257	Q62257 mus musculu
44	43	43.9	423	17	Q91900	Q91900 aeropyrum p
45	43	43.9	477	5	Q8SR34	Q8SR34 encephalit
			484	10	Q91U12	Q91U12 blastocyst

## ALIGNMENTS

RESULT 1	ID	Q9D695	PRELIMINARY:	PRT:	360 AA.
AC	Q9D695	01-JUN-2001 (TREMBLrel. 17, Created)			
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)				
DE	01-JUN-2002 (TREMBLrel. 21, Last annotation update)				
DE	4631416M05RIK protein (Megs1n).				
GN	SERPINF7 OR 4631416M05RIK.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_Taxid:10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	SRRATN-C57BL/6J; TISSUE=SKIN;				
RX	MEDLINE=21085660; PubMed=11217851;				
RA	Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,				
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,				
RA	Alzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,				
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,				
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,				
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,				
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,				
RA	Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,				
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,				
RA	Blake J., Boffelli D., Bojunga N., Carinici P., de Bonaldo M.F.,				
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,				
RA	Gastlreich S., Hill D., Hofmann M., Hume D.A., Kamiya K., Lee N.H.,				
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,				
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,				
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,				
RA	Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,				
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,				
RA	Hayashizaki Y.,				
RT	"Functional annotation of a full-length mouse cDNA collection.";				
RL	Nature 409:685-690(2001).				
RN	[2]				
RP	SEQUENCE FROM N.A.				

```

RX MEDLINE-21368006: PubMed-11473647;
RA Nangaku M., Miyata T., Suzuki D., Umezono T., Hashimoto T., Wada T.,
RA Yagi M., Nagano N., Inagi R., Kurokawa K.;
RT "Cloning of rodent meginin revealed its up-regulation in
RT mesangio proliferative nephritis.";
RL Kidney Int. 60:641-652(2001).
DR EMBL: AK014524; BAB29410.1; -.
DR EMBL: AF105328; AAL16768.1; -.
DR HSSP: P05121; 1A7C.
DR MGD: MGI:2151053; Serpinb7.
DR InterPro: IPR000215; Serpin.
DR Pfam: PF00079; serpin.1.
DR SMART: SM00093; SERPIN; 1.
DR PROSITE: PS00284; SERPIN; UNKNOWN_1.
SQ SEQUENCE 380 AA; 43050 MW; C9240272BCF9CFA CRC64;

Query Match
Best Local Similarity 84.7%; Score 83; DB 11; Length 380;
Matches 16; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 ADLGIASGGRLYISRMKG 20
DB 305 ADLGIASGGRLYYSKLMHK 324

RESULT 2
Q920J5 PRELIMINARY; PRT; 380 AA.
AC Q920J5;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Meginin.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-21368006: PubMed-11473647;
RA Nangaku M., Miyata T., Suzuki D., Umezono T., Hashimoto T., Wada T.,
RA Yagi M., Nagano N., Inagi R., Kurokawa K.;
RT "Cloning of rodent meginin revealed its up-regulation in
RT mesangio proliferative nephritis.";
RL Kidney Int. 60:641-652(2001).
CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.
DR EMBL: AF105329; AAL16769.1; -.
DR InterPro: IPR000215; Serpin.
DR Pfam: PF00079; serpin.1.
DR PROSITE: PS00284; SERPIN; UNKNOWN_1.
KW Serpin.
SQ SEQUENCE 380 AA; 42821 MW; D8076CA8EE2C2FBC CRC64;

Query Match
Best Local Similarity 84.7%; Score 83; DB 11; Length 380;
Matches 16; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 ADLGIASGGRLYISRMKG 20
DB 305 ADLGIASGGRLYYSKLMHK 324

RESULT 3
Q63547 PRELIMINARY; PRT; 191 AA.
AC Q63547;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ZG-21P.
GN ZG-21.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WISTAR: TISSUE-PANCREAS;
RA Cronshagen U., Chen C., Kern H.F.;
RT "A novel protein expressed exclusively in pancreas is proposed to be a
RT serpin.";
RL Submitted (MAR-1994) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.
DR EMBL: Z30585; CAAB3060.1; -.
DR HSSP: P01008; 1A7H.
DR InterPro: IPR000215; Serpin.
DR Pfam: PF00079; serpin.1.
DR SMART: SM00093; SERPIN; 1.
DR PROSITE: PS00284; SERPIN; UNKNOWN_1.
KW Serpin.
SQ SEQUENCE 191 AA; 21825 MW; 9EC312FEDBD169320 CRC64;

Query Match
Best Local Similarity 55.1%; Score 54; DB 11; Length 191;
Matches 11; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 2 DLGIASGGRLYISRMKG 20
DB 102 DLGITDSSSELYVSRAK 120

RESULT 4
Q9N0X3 PRELIMINARY; PRT; 615 AA.
AC Q9N0X3;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CDNA FLJ11078 fis, clone PLACE1005102, weakly similar to RING CANAL
DE protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-PLACENTA;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;
RT "NEO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AK001940; BAA91990.1; -.
DR HSSP: Q05516; 1CS3.
DR InterPro: IPR000210; BTB_POZ.
DR Pfam: PF00651; BTB; 1.
DR Pfam: PF01344; Kelch; 5.
DR SMART: SM00225; BTB; 1.
DR PROSITE: PS50097; BTB; 1.
SQ SEQUENCE 615 AA; 68139 MW; 1583474AF5A45E8D CRC64;

Query Match
Best Local Similarity 51.0%; Score 50; DB 4; Length 615;
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 SGIASGGRLYIS 15
DB 454 AGAASGGRLYIS 465

RESULT 5
Q8TAP0 PRELIMINARY; PRT; 615 AA.
AC Q8TAP0;

```

```

DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 68.2 kDa protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC026319; AAR26319.1; -.
KW Hypothetical protein;
SQ SEQUENCE 615 AA; 68171 MW; 159EDFDE80EE98D CRC64;

Query Match          51.0%; Score 50; DB 4; Length 615;
Best Local Similarity 83.3%; Pred. No. 9.9;
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 4 SG1ASGRLYIS 15
Db 454 AGAASGRLYIS 465

RESULT 6
ID Q9D7P9 PRELIMINARY; PRT; 423 AA.
AC Q9D7P9;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE 220003F07RIK protein.
GN 220003F07RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TONGUE;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H. A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirml L. M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barish G.,
RA Blake J., Boileau D., Bojunga N., Carninci P., de Bonaldo M. F.,
RA Brownstein M. J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D. A., Kamiya M., Lee N. H.,
RA Lyons P., Marchionni L., Mashima J., Mazarrelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K. F.,
RA Suzuki H., Toyooka K., Wang K. H., Wetz C., Whitaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC - SIMILARITY: BELONGS TO THE SERPIN FAMILY.
DR EMBL: AK009018; BAB26028.1; -.
DR HSSP: P05619; 1HLE.
DR MGD: MGI:1919119; 2300003F07RIK.
DR InterPro: IPR000215; Serpin.
DR Pfam: PF00079; serpin.1.
DR SMART: SM00093; SERPIN_1.
DR PROSITE: PS00284; SERPIN; 1.
KW Serpin.
SQ SEQUENCE 423 AA; 47834 MW; 5A22BE2FE51B6120 CRC64;

```

```

Query Match          49.0%; Score 48; DB 11; Length 423;
Best Local Similarity 45.0%; Pred. NO. 14;
Matches 9; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Oy 1 ADLSGASGRLYISRMKG 20
Db 346 ADLGIKSPNLYIKIVKH 365

RESULT 7
ID Q9RM57 PRELIMINARY; PRT; 591 AA.
AC Q9RM57;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Serine protease, subtilase family.
GN DR0812.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J. A., Heidelberg J. F., Hickey E. K., Peterson J. D.,
RA Dodson R. J., Haft D. H., Gwinn M. L., Nelson W. C., Richardson D. L.,
RA Moffat K. S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J. J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K. S., Aravind L., Daly M. J., Minton K. W., Fleischmann R. D.,
RA Ketchum K. A., Nelson K. E., Salzberg S., Smith H. O., Venter J. C.,
RA Fraser C. M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577(1999).
DR EMBL: AF001935; AAF10389.1; -.
DR HSSP: Q99405; IMPT.
DR TIGR: DR0812; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR InterPro: IPR003137; PA.
DR InterPro: IPR000209; Peptidase_s8.
DR Pfam: PF02225; PA; 1.
DR Pfam: PF00082; Peptidase_S8; 2.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR PROSITE: PS00138; SUBTILASE_SER; UNKNOWN_1.
KW Protease; Complete proteome.
SQ SEQUENCE 591 AA; 59158 MW; 2EBAD3440D1FA5E4 CRC64;

Query Match          49.0%; Score 48; DB 16; Length 591;
Best Local Similarity 47.1%; Pred. NO. 20;
Matches 8; Conservative 47; 6; Mismatches 3; Indels 0; Gaps 0;

Oy 3 LSG1ASGRLYISRMKG 19
Db 255 VGVGASGVNLYMARVLG 271

RESULT 8
ID Q9MZ16 PRELIMINARY; PRT; 327 AA.
AC Q9MZ16;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE GRIH receptor.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RA MEDLINE-20267228; PubMed-10809231;
RX Cui J., Smith R.G., Mount G.R., Lo J.L., Yu J., Walsh T.F.,
RA Singh S.B., Devita R.J., Goulet M.T., Schaeffer J.M., Cheng K.;
RT "Identification of Phe313 of the gonadotropin-releasing hormone (GNRH)
RT receptor as a site critical for the binding of nonpeptide GnRH
antagonists."
RL Mol. Endocrinol. 14:671-681(2000).
DR EMBL; AF206513; AAF87097.1;
DR InterPro: IPR000276; GPCR_Rhodopsn.
DR Pfam: PF00001; 7tm.1.1.
DR PRINTS: PR00237; GPCRHOOPS.
DR PROSITE: PS00237; G_PROTEIN_RECEP_F1_1; UNKNOWN.1.
DR PROSITE: PS50262; G_PROTEIN_RECEP_F1_2; 1.
DR Receptor.
SQ SEQUENCE 327 AA; 37655 MW; 778A82A9D25D3371 CRC64;

Query Match 46.9%; Score 46; DB 6; Length 327;
Best Local Similarity 62.5%; Pred. No. 22;
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 3 LSGIAGSGRLYISRM 18
DB 165 LSGIFAGPOLXIFRMI 180
|||||:|:|||||:|

RESULT 9
ID 008804 PRELIMINARY; PRT; 377 AA.
AC 008804;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Serine proteinase inhibitor NK13.
GN SPI12.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C57BL;
RA MEDLINE-97326124; PubMed-9182575;
RX Sun J., Ooms L., Bird C.H., Sutton V.R., Trapani J.A., Bird P.I.;
RT "A new family of 10 murine ovalbumin serpins includes two homologs of
RT proteinase inhibitor 8 and two homologs of the granzyme B inhibitor
RT (proteinase inhibitor 9).";
RL J. Biol. Chem. 272:15434-15441(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-C57BL;
RA Kalsnerman D., Knaggs S., Scarff K.L., Gillard A., Mirza G., Cadman M.,
RA McKee R., Denny P., Cooley J., Benarafa C., Remold-O'Donnell E.,
RA Ragoussis J., Bird P.I.;
RT "Comparison of human serpin gene 6p25 with murine chromosome 13 reveals
RT a greatly expanded Ov-serpin gene repertoire in the mouse.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.
DR EMBL; U96707; AAB57819.2;
DR GMD; MGI:894688; Sp112.
DR InterPro: IPR000215; Serpin.
DR Pfam; PF00079; serpin.1.
DR SMART; SM00093; SERPIN; 1.
DR PROSITE; PS00284; SERPIN; 1.
DR Serpin.
KW SEQUENCE 377 AA; 42535 MW; 9660B8F3A41EE47A CRC64;

Query Match 46.9%; Score 46; DB 11; Length 377;
Best Local Similarity 50.0%; Pred. No. 26;
Matches 10; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

OY 1 ADLSGIAGSGRLYISRM 20
DB 11 ADLSGIAGSGRLYISRM 20
|||||:|:|||||:|

```

```

DB 302 ADFSGIAGSGRLYISRM 321

RESULT 10
ID 0967L9 PRELIMINARY; PRT; 382 AA.
AC 0967L9;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Serpin (Fragment).
GN SPI.
OS Schistosoma japonicum (Blood fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigoidae;
OC Schistosomatidae; Schistosomatidae; Schistosoma.
OX NCBI_TaxID=6182;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CHINESE;
RA Yan Y.T., Liu S.X.;
RT "Serpins Sequence of Schistosoma japonicum.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF317731; AAC45932.1;
DR InterPro: IPR00215; Serpin.
DR Pfam; PF00079; serpin.1.
DR SMART; SM00093; SERPIN; 1.
DR NON_TER 1
FT 1
SQ SEQUENCE 382 AA; 43162 MW; B6B816CFA868D1E7 CRC64;

Query Match 46.9%; Score 46; DB 5; Length 382;
Best Local Similarity 60.0%; Pred. No. 27;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 ADLSGIAGSGRLYIS 15
DB 302 ADLSGIAGSGRLYIS 316
|||||:|:|||||:|

RESULT 11
ID 0967L9 PRELIMINARY; PRT; 400 AA.
AC 0967L9;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 20, Last annotation update)
DE Serine protease inhibitor serpin.
GN SPI.
OS Schistosoma japonicum (Blood fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigoidae;
OC Schistosomatidae; Schistosomatidae; Schistosoma.
OX NCBI_TaxID=6182;
RN [1]
RP SEQUENCE FROM N.A.
RA Yan Y.T., Liu S.X., Song G.C.;
RT "Full-length of Serpin Sequence of Schistosoma japonicum.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.
DR EMBL; AF380366; AAK57435.1;
DR InterPro: IPR000215; Serpin.
DR Pfam; PF00079; serpin.1.
DR Protease; Serpin.
KW SEQUENCE 400 AA; 45207 MW; 36BB918B393689BF CRC64;

Query Match 46.9%; Score 46; DB 5; Length 400;
Best Local Similarity 60.0%; Pred. No. 28;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 ADLSGIAGSGRLYIS 15
DB 320 ADLSGIAGSGRLYIS 334
|||||:|:|||||:|

```

RESULT 12



Q9NH65  
ID Q9NH65 PRELIMINARY; PRT; 373 AA.  
AC Q9NH65;  
DT 01-OCT-2000 (TREMUREL. 15, Created)  
DT 01-OCT-2000 (TREMUREL. 15, Last sequence update)  
DT 01-JUN-2002 (TREMUREL. 21, Last annotation update)  
DE Serine proteinase inhibitor.  
OS Trichinella spiralis.  
OC Eukaryota; Metazoa; Nematoda; Enopleae; Trichocephalida;  
OC Trichinelidae; Trichinella.  
OX NCBI\_TaxID=6334;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ISS413;  
RA Nagano I., Wu Z., Takahashi Y.;  
RT "Molecular cloning and expression of serine proteinase inhibitor from  
Trichinella spiralis."  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.  
DR EMBL: AF231948; AAF63473.1; -.  
DR HSSP: P01008; IATH.  
DR InterPro: IPR000215; Serpin.  
DR Pfam: PF00079; Serpin.1.  
DR SMART: SM00093; SERPIN.1.  
DR PROSITE: PS00284; SERPIN.1.  
KM Serpin.  
SQ SEQUENCE 373 AA; 42414 MW; 78CAACCDAC06BCF5 CRC64;  
Query Match 45.98; Score 45; DB 5; Length 373;  
Best Local Similarity 50.0%; Pred. No. 38;  
Matches 10; Conservative 4; Mismatches 6; Indels 0; Gaps 0;  
QY 1 ADLSGIASGRLYSRMGK 20  
Db 293 ADLSGICVKEKLYSDVHK 312  
RESULT 13  
Q26509 PRELIMINARY; PRT; 385 AA.  
AC Q26509;  
DT 01-NOV-1996 (TREMUREL. 01, Created)  
DT 01-NOV-1996 (TREMUREL. 01, Last sequence update)  
DT 01-OCT-2001 (TREMUREL. 18, Last annotation update)  
DE Serpin (Fragment).  
GN SPI.  
OS Schistosoma japonicum (Blood fluke).  
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigoidida;  
OC Schistosomatoidea; Schistosomatidae; Schistosoma.  
OX NCBI\_TaxID=6182;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ogunlunde J.O., Raehid A.A., Licate L.S., Liu S., Blanton R.E.;  
RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 280-385 FROM N.A.  
RC STRAIN-PHILIPPINE;  
RA Hooker C.W., Brindley P.J.;  
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.  
DR EMBL: U11023; AAA50230.1; -.  
DR EMBL: AF043997; AAC02994.1; -.  
DR InterPro: IPR000215; Serpin.  
DR Pfam: PF00079; Serpin.1.  
DR SMART: SM00093; SERPIN.1.  
KM Serpin.  
FT NON\_TER  
SQ SEQUENCE 385 AA; 43439 MW; 652E15AD670C4719 CRC64;  
Query Match 45.98; Score 45; DB 5; Length 385;  
Best Local Similarity 64.3%; Pred. No. 40;  
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 ADLSGIASGRLYI 14  
Db 305 ADLSGISKNKLV 318  
RESULT 14  
Q65309 PRELIMINARY; PRT; 942 AA.  
AC Q65309;  
DT 01-AUG-1998 (TREMUREL. 07, Created)  
DT 01-AUG-1998 (TREMUREL. 07, Last sequence update)  
DT 01-JUN-2002 (TREMUREL. 21, Last annotation update)  
DE Hypothetical 104.5 kDa protein.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPRAIN-CV. WS;  
RA Sanders P.M., Bul A.O., Weterings K., McIntire K.N., Hsu Y.C.,  
Lee P.Y., Truong M.T., Beals T.B., Goldberg R.B.;  
RT "Another Development Defects in Arabidopsis thaliana Male-Sterile  
Mutants."  
RL Sex. Plant Reprod. 11:297-322(1999).  
DR EMBL: AF060248; AAC97105.1; -.  
DR InterPro: IPR001159; DS\_RBD.  
DR InterPro: IPR000051; SAM\_bind.  
DR Pfam: PF00035; dsrm.1.  
KM Hypothetical protein.  
SQ SEQUENCE 942 AA; 104497 MW; E09B964161CA8F69 CRC64;  
Query Match 45.98; Score 45; DB 10; Length 942;  
Best Local Similarity 45.0%; Pred. No. 11e+02;  
Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;  
QY 1 ADLSGIASGRLYSRMGK 20  
Db 250 AERLGLKGNQVMISRMFGK 269  
RESULT 15  
Q9C508 PRELIMINARY; PRT; 942 AA.  
AC Q9C508;  
DT 01-JUN-2001 (TREMUREL. 17, Created)  
DT 01-JUN-2001 (TREMUREL. 17, Last sequence update)  
DT 01-JUN-2002 (TREMUREL. 21, Last annotation update)  
DE CORYMBOSA2.  
GN CRM2.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Suzuki M., Takahashi T., Kameda Y.;  
RL "CORYMBOSA2 (CRM2), a gene required for inflorescence development in  
Arabidopsis thaliana."  
RT Arabidopsis thaliana.  
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF327068; AKI16435.1; -.  
DR InterPro: IPR001159; DS\_RBD.  
DR InterPro: IPR000051; SAM\_bind.  
DR Pfam: PF00035; dsrm.1.  
SQ SEQUENCE 942 AA; 104454 MW; 7DBB5222C0D28779 CRC64;  
Query Match 45.98; Score 45; DB 10; Length 942;  
Best Local Similarity 45.0%; Pred. No. 1.1e+02;  
Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;  
QY 1 ADLSGIASGRLYSRMGK 20

DB 250 AERLGLKDGNOVMISRMFGK 269

Search completed: July 11, 2003, 11:59:02  
Job time : 10.7318 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:42:16 ; Search time 2.86902 Seconds  
(without alignments)  
232.223 Million cell updates/sec

Title: US-10-091-442-7

Perfect score: 26

Sequence: 1 LYDAK 5

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002:\*

- 1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*
- 2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*
- 3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*
- 4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*
- 5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*
- 6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*
- 7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*
- 8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*
- 9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*
- 10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*
- 11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*
- 12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*
- 13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*
- 14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*
- 15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*
- 16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*
- 17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*
- 18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*
- 19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*
- 20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*
- 21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*
- 22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*
- 23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	26	100.0	5	AAK57113	Human megakaryocyte
2	26	100.0	58	AAAB16557	Bacteriophage 44AH
3	26	100.0	58	ABH81848	Staphylococcus aur
4	26	100.0	69	ABP22986	Streptococcus poly
5	26	100.0	70	ABP229293	Streptococcus poly
6	26	100.0	118	AAAB51492	Human secreted pro
7	26	100.0	118	AAAB51494	Human secreted pro
8	26	100.0	151	AAAG28475	Arabidopsis thalia
9	26	100.0	155	AAAG28474	Arabidopsis thalia
10	26	100.0	212	ABG24830	Novel human diagno

11	26	100.0	253	23	ABB49942
12	26	100.0	263	21	AAAG25581
13	26	100.0	263	21	AAAG37820
14	26	100.0	297	21	AAAG25580
15	26	100.0	297	21	AAAG37819
16	26	100.0	319	22	AAAB79071
17	26	100.0	357	21	AAAG25579
18	26	100.0	357	21	AAAG37818
19	26	100.0	357	23	AAAB93723
20	26	100.0	380	15	AAAB48379
21	26	100.0	380	21	AAAB08254
22	26	100.0	380	21	AAAB24142
23	26	100.0	380	22	AAAB83075
24	26	100.0	413	22	AAAG90403
25	26	100.0	427	22	AAAB79070
26	26	100.0	491	22	AAAB10377
27	26	100.0	588	22	AAAB15599
28	26	100.0	630	21	AAAB57586
29	26	100.0	745	22	AAAB36616
30	26	100.0	969	22	AAAB71685
31	26	100.0	1014	22	AAAG90902
32	26	100.0	1043	18	AAAB18306
33	26	100.0	1043	19	AAAB56569
34	26	100.0	1213	22	AAAB71439
35	26	100.0	1585	22	AAAG60675
36	26	100.0	1585	22	AAAB10815
37	26	100.0	1588	22	AAAG06618
38	26	100.0	1598	22	AAAG09655
39	26	100.0	1647	22	AAAG04001
40	26	100.0	1844	21	AAAB18250
41	26	100.0	2030	22	AAAB64300
42	26	100.0	2424	22	AAAB68256
43	26	100.0	2918	22	AAAG27218
44	26	100.0	73	22	AAAB5183
45	24	92.3	88	19	AAAB77753

#### ALIGNMENTS

RESULT 1	AAK57113	standard; peptide; 5 AA.
ID	AAK57113	
XX	AAK57113:	
AC	16-AUG-1994 (first entry)	
XX		
DT		
XX		
DE	Human megakaryocyte differentiation factor peptide 7.	
XX		
KW	Human megakaryocyte differentiation factor; MDF; thrombopoietin;	
KW	haematopoietic stimulating factor; thrombocytopoenta; platelet;	
KW	bone marrow transplantation; cancer chemotherapy.	
XX		
OS	Homo sapiens.	
XX		
PN	EP583884-A.	
XX		
PD	23-FEB-1994.	
XX		
PF	19-JUL-1993: 93EP-0305654.	
XX		
PR	17-JUL-1992: 92JP-0212305.	
XX		
PR	04-MAR-1993: 93JP-0067339.	
XX		
PA	(SUNR) SUNTORY LTD.	
XX	(TSUT) TSUTIMOTO M.	
PI	Ishida N, Iwasa F, Kurihara T, Miura K, Nakazato H;	
PI	Tsujimoto M, Tsunoka N, Yamaguchi N, Yamachi K;	
XX		
DR	WPI: 1994-058782/08.	
XX		

Listeria monocytog  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Corynebacterium gl  
Arabidopsis thalia  
Arabidopsis thalia  
Hebically activ  
Human megakaryocyt  
Human megakaryocyt  
Human megakaryocyt  
Human megakaryocyt  
Human megakaryocyt  
C glutamicum prote  
Corynebacterium gl  
Salmonella typhi (   
Novel human diagno  
Human papillomavir  
Staphylococcus aur  
Drosophila melanog  
C glutamicum prote  
Photobacterium lum  
Toxin TccC, encode  
Drosophila melanog  
Novel human diagno  
Novel human diagno  
Novel human diagno  
Novel human diagno  
Plasmodium falcipa  
Drosophila melanog  
Novel human diagno  
Human immune/haema  
pyruvate oxidase p

PT New megakaryocyte differentiation factor - isolated from human  
 PT epidermoid carcinoma cells, used to treat conditions involving a  
 decrease in platelets  
 XX  
 PS Claim 1: Page 20; 47pp; English.  
 XX  
 CC Human MDF (see AAR48379) can be isolated from a culture of human  
 CC epidermoid carcinoma A431 cells in protein-free medium. The MDF  
 CC stimulates differentiation of megakaryocytes from myeloid cells  
 CC in the presence of IL-3. The MDF acts in vivo as a thrombopoietin  
 CC making it useful for treatment of diseases involving a decrease  
 CC in platelet number (esp. thrombocytopenia) such as occurs in bone  
 CC marrow transplantation and in chemotherapy. MDF has mol.wt.  
 CC 55-57kD by SDS-PAGE, exhibits an isoelectric point of 6.0-7.0 and  
 CC contains an amino acid sequence comprising at least one of the  
 CC sequences AAR57107-R57115.  
 CC  
 SQ Sequence 5 AA;  
 Query Match 100.0%; Score 26; DB 15; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LYDAK 5  
 DB 1 LYDAK 5  
 RESULT 2  
 AAB16557  
 ID AAB16557 standard; Protein; 58 AA.  
 XX  
 AC AAB16557;  
 XX  
 DT 27-OCT-2000 (first entry)  
 XX  
 DE Bacteriophage 44AHJD protein sequence 44AHJDORF025.  
 XX  
 KW Bacteriophage; antimicrobial; genome; identification; antibacterial;  
 KW bacterial growth inhibition; bacterial infection.  
 XX  
 OS Bacteriophage 44AHJD.  
 XX  
 PN WO200032825-A2.  
 PD  
 XX  
 PD 08-JUN-2000.  
 XX  
 PF 03-DEC-1999; 99WO-IB02040.  
 XX  
 PR 03-DEC-1998; 98US-0110992.  
 PR 03-JUN-1999; 99US-0326144.  
 PR 28-SEP-1999; 99US-0407804.  
 PR 30-SEP-1999; 99US-0157218.  
 PR 01-DEC-1998; 99US-0168777.  
 PR 02-DEC-1999; 99US-0434252.  
 XX  
 PA (PHAG-) PHAGETECH INC.  
 XX  
 PI Pelletier J, Gros P, Dubow M;  
 XX  
 DR WPI: 2000-412361/35.  
 DR N-PSDB: AAA69042.  
 XX  
 PT Identifying a bacteriophage coding region for treating bacterial  
 PT infections comprises identifying a nucleic acid encoding a product that  
 PT inhibits bacteria when a bacteriophage infects a bacterium -  
 XX  
 PS Example 9; Page 278; 456pp; English.  
 PS  
 CC The present invention describes a method for identifying a bacteriophage  
 CC coding region encoding a product active on an essential bacterial  
 CC target. The method comprises identifying a nucleic acid sequence encoding  
 CC a gene product that provides a bacteria-inhibiting function when an

CC uncharacterised bacteriophage infects a pathogenic bacterium. The  
 CC compound active on a target of a bacteriophage inhibitor protein in a  
 CC bacteria is used to treat or prevent a bacterial infection in an animal.  
 CC AAA68243 to AAA69442 and AAB16523 to AAB16954 represent bacteriophage  
 CC nucleotide and protein sequences which are used in the exemplification of  
 CC the present invention.  
 CC  
 SQ Sequence 58 AA;  
 Query Match 100.0%; Score 26; DB 21; Length 58;  
 Best Local Similarity 100.0%; Pred. No. 63;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LYDAK 5  
 DB 29 LYDAK 33  
 RESULT 3  
 ABB81848  
 ID ABB81848 standard; Protein; 58 AA.  
 XX  
 AC ABB81848;  
 XX  
 DT 20-SEP-2002 (first entry)  
 XX  
 DE Staphylococcus aureus 44AHJD ORF 25 amino acid sequence.  
 XX  
 KW STAU-R2; drug screening; antibacterial; bacterial infection;  
 KW 44AHJD ORF 25.  
 XX  
 OS Staphylococcus aureus.  
 XX  
 PN WO200244718-A2.  
 PD  
 XX  
 PD 06-JUN-2002.  
 XX  
 PF 30-NOV-2001; 2001WO-CA01754.  
 XX  
 PR 30-NOV-2000; 2000US-0727892.  
 PR 20-JUN-2001; 2001US-0885561.  
 XX  
 PA (PHAG-) PHAGETECH INC.  
 XX  
 PI Pelletier J, Gros P, Dubow M;  
 XX  
 DR WPI: 2002-566592/60.  
 DR N-PSDB: ABB86788.  
 XX  
 PT Use of a Staphylococcus aureus (STAU-R2) polypeptide or composition  
 PT comprising STAU-R2 and polypeptide encoded by bacteriophage open  
 PT reading frame that specifically interacts with STAU-R2, for identifying  
 PT a compound active on STAU-R2 -  
 XX  
 PS Claim 2; Fig 2; 131pp; English.  
 PS  
 CC The invention relates to a novel staphylococcus aureus polypeptide  
 CC (STAU-R2), and it's use in drug screening assays. The polypeptide of the  
 CC invention has antibacterial activity. The compound active on STAU-R2 is  
 CC useful for the manufacture of an antibacterial agent or for the  
 CC manufacture for medicament for treating or preventing a bacterial  
 CC infection. The sequence represents the amino acid sequence of S. aureus  
 CC 44AHJD ORF 25.  
 CC  
 SQ Sequence 58 AA;  
 Query Match 100.0%; Score 26; DB 23; Length 58;  
 Best Local Similarity 100.0%; Pred. No. 63;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LYDAK 5  
 DB 29 LYDAK 33

## RESULT 4

ABP29986 standard; Protein: 69 AA.

ABP29986;

02-JUL-2002 (first entry)

Streptococcus polypeptide SEQ ID NO 9148.

Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;

group A streptococcus; Streptococcus pyogenes; antibacterial;

antiinflammatory; infection; vaccine; meningitis; gene therapy.

Streptococcus pyogenes.

MO200234771-A2.

02-MAY-2002.

29-OCT-2001; 2001WO-GB04789.

27-OCT-2000; 2000GB-0026333.

24-NOV-2000; 2000GB-0028727.

07-MAR-2001; 2001GB-0005640.

(CHIR-) CHIRON SPA.

(GENO-) INST GENOMIC RES.

Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;

Tetelin H;

WPI: 2002-352536/38.

N-PSDB: ABN70617.

New Streptococcus protein for the treatment or prevention of infection

or disease caused by Streptococcus bacteria, such as meningitis, and

for detecting a compound that binds to the protein -

Claim 1; Page 4037; 4525pp; English.

The invention relates to a protein (ABP25413-ABP30895) from group B

(Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in

the specification. The proteins have antibacterial and antiinflammatory

activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and

antibodies that bind (I) are used in the manufacture of medicaments for

the treatment or prevention of infection or disease caused by

Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.

Nucleic acids encoding (I) are used to detect Streptococcus in a

biological sample. (I) is used to determine whether a compound binds to

(I). A composition comprising (I) or a nucleic acid encoding (I), may be

used as a vaccine or diagnostic composition. The disease caused by

Streptococcus that is prevented or treated may be meningitis. Nucleic

acid encoding (I) may be used to recombinantly produce (I) and may be

used in gene therapy. Antibodies to (I) are used for affinity

chromatography, immunoassays, and distinguishing/identifying

Streptococcus proteins.

Sequence 69 AA:

Query Match 100.0%; Score 26; DB 23; Length 69;

Best Local Similarity 100.0%; Pred. No. 76;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5

DB 32 LYDAK 36

RESULT 5

ABP29293 standard; Protein: 70 AA.

ABP29293;

02-JUL-2002 (first entry)

Streptococcus polypeptide SEQ ID NO 7762.

Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;

group A streptococcus; Streptococcus pyogenes; antibacterial;

antiinflammatory; infection; vaccine; meningitis; gene therapy.

Streptococcus pyogenes.

MO200234771-A2.

02-MAY-2002.

29-OCT-2001; 2001WO-GB04789.

27-OCT-2000; 2000GB-0026333.

24-NOV-2000; 2000GB-0028727.

07-MAR-2001; 2001GB-0005640.

(CHIR-) CHIRON SPA.

(GENO-) INST GENOMIC RES.

Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;

Tetelin H;

WPI: 2002-352536/38.

N-PSDB: ABN69924.

New Streptococcus protein for the treatment or prevention of infection

or disease caused by Streptococcus bacteria, such as meningitis, and

for detecting a compound that binds to the protein -

Claim 1; Page 3912; 4525pp; English.

The invention relates to a protein (ABP25413-ABP30895) from group B

(Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in

the specification. The proteins have antibacterial and antiinflammatory

activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and

antibodies that bind (I) are used in the manufacture of medicaments for

the treatment or prevention of infection or disease caused by

Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.

Nucleic acids encoding (I) are used to detect Streptococcus in a

biological sample. (I) is used to determine whether a compound binds to

(I). A composition comprising (I) or a nucleic acid encoding (I), may be

used as a vaccine or diagnostic composition. The disease caused by

Streptococcus that is prevented or treated may be meningitis. Nucleic

acid encoding (I) may be used to recombinantly produce (I) and may be

used in gene therapy. Antibodies to (I) are used for affinity

chromatography, immunoassays, and distinguishing/identifying

Streptococcus proteins.

Sequence 70 AA:

Query Match 100.0%; Score 26; DB 23; Length 70;

Best Local Similarity 100.0%; Pred. No. 77;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5

DB 33 LYDAK 37

RESULT 6

AAB51492 standard; Protein: 118 AA.

AAB51492

XX

AC	AAB51492;
XX	
DT	16-FEB-2001 (first entry)
DE	Human secreted protein BLAST search protein SEQ ID NO: 169.
XX	
KW	Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KM	antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
KV	vulnerary; anticoagulant; antibacterial; antifungal; antiparasitic;
KX	cardiac; gene therapy; cancer; immune disorder cardiovascular disorder;
XX	neurological disease; infection; human; secreted protein.
OS	Homo sapiens.
PN	WO200058495-A1.
PD	05-OCT-2000.
PF	23-MAR-2000; 2000MO-US07661.
PR	26-MAR-1999; 99US-0126504.
PR	07-JAN-2000; 2000US-0174847.
PA	(HUMA-) HUMAN GENOME SCI INC.
PJ	Rosen CA, Ruden SM, Komatsoulis G;
DR	WPI; 2000-611720/58.
PT	New nucleic acid molecules encoding 45 human secreted proteins for
PT	diagnosing, preventing, treating or ameliorating medical conditions and
PS	used as food additives or preservatives -
XX	
PS	Disclosure; Page 400; 410pp; English.
CC	The invention relates to the isolation of genes AAc93310-C93354 encoding
CC	the human secreted proteins AAB51380-B51423. The genes and proteins are
CC	useful for preventing, ameliorating or treating medical conditions, e.g.
CC	by protein or gene therapy. The genes are isolated from a range of human
CC	tissues disclosed in the specification. The nucleic acids, proteins,
CC	antibodies and (ant)agonists are useful in the diagnosis, treatment and
CC	prevention of: (a) cancer, e.g. breast and ovarian cancer, and other
CC	cancers of the adrenal gland, bone, bone marrow, breast,
CC	gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders
CC	e.g. Addison's disease, allergies, autoimmune haemolytic anaemia,
CC	autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple
CC	sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC	cardiovascular disorders such as myocardial ischaemias; (d) wound
CC	healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
CC	and (f) infectious diseases such as viral, bacterial, fungal and
CC	parasitic infections. The present sequence is a protein isolated in the
CC	present invention.
XX	
SQ	Sequence 118 AA:
QY	
DB	
	Query Match 100.0%; Score 26; DB 21; Length 118;
	Best Local Similarity 100.0%; Pred. No. 1,4e+02;
	Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0
	1 LYDAK 5
	77 LYDAK 81
RESULT 7	
ID	AAB51494
ID	AAB51494 standard; Protein; 118 AA.
XX	
AC	AAB51494;
XX	
DT	16-FEB-2001 (first entry)
XX	
DE	Human secreted protein BLAST search protein SEQ ID NO: 171.

XX	Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KM	antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
KM	vulnerable; anticonvulsant; antibacterial; antitumoral; antiparasitic;
KW	cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX	neurological disease; infection; human; secreted protein.
XX	
OS	Homo sapiens.
XX	
PN	WO200058495-A1.
XX	
PD	05-OCT-2000.
XX	
PF	23-MAR-2000; 2000WO-US07661.
XX	
PR	26-MAR-1999; 99US-0126504.
PR	07-JAN-2000; 2000US-0174847.
XX	
PA	(HUMA-) HUMAN GENOME SCL INC.
XX	
PI	Rosen CA, Ruben SM, Komatsoulis G;
XX	
DR	WPI, 2000-611720/58.
XX	
PS	Disclosure; Page 401; 410pp; English.
XX	
CC	The invention relates to the isolation of gene AHC93310-C93354 encoding
CC	the human secreted proteins AHB51380-B51423. This sequence represents a
CC	peptide fragment homologous to the protein encoded by the gene isolated
CC	in the present invention. The sequence is a search result from a BLASTX
CC	homology search. The genes and proteins are useful for preventing,
CC	ameliorating or treating medical conditions, e.g. by protein or gene
CC	therapy. The genes are isolated from a range of human tissues disclosed
CC	in the specification. The nucleic acids, proteins, antibodies and
CC	(ant)agonists are useful in the diagnosis, treatment and prevention of:
CC	(a) cancer, e.g. breast and ovarian cancer, and other cancers of the
CC	adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
CC	lung, or urogenital; (b) immune disorders e.g. Addison's disease,
CC	allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC	diabetes mellitus, Cronh's disease, multiple sclerosis, rheumatoid
CC	arthritis and ulcerative colitis; (c) cardiovascular disorders such
CC	as myocardial ischaemias; (d) wound healing; (e) neurological diseases
CC	e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such
CC	as viral, bacterial, fungal and parasitic infections.
XX	
SO	Sequence 118 AA;
Query Match	100.0%; Score 26; DB 21; Length 118;
Best Local Similarity	100.0%; Pred. No. 1.4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	1 LYDAK 5
Db	77 LYDAK 81
RESULT 8	
ID	AA628475
AA628475	standard; Protein: 151 AA.
XX	
AC	AA628475;
XX	
DT	17-OCT-2000 (first entry)
XX	
XX	
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 33705.
KM	Protein identification; signal transduction pathway; metabolic pathway;
KW	hybridisation assay; genetic mapping; gene expression control; promoter;
XX	termination sequence.

Fri Jul 11 15:11:17 2003

us-10-091-442-7.rag

Page 5

```
XX Arbidopsis thaliana.
OS
XX EPI033405-A2.
XX
XX
PD 06-SEP-2000.
PF
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 04-MAY-1999; 99US-0132048.
PR 05-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135153.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145092.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
```

PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155859.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
  
Query Match 100.0%; Score 26; DB 21; Length 151;  
Best Local Similarity 100.0%; Pred. No. 1,8e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 LYDAK 5  
DB 2 LYDAK 6  
  
RESULT 9  
AAG28474  
ID AAG28474 standard; Protein; 155 AA.  
XX AAG28474;  
AC  
XX  
DF 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 33704.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX PD 06-SEP-2000.  
XX

PR 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 27-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.



```

PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151338.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.

```

```

PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161820.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

Query Match 100.0%; Score 26; DB 21; Length 155;
Best Local Similarity 100.0%; Pred. No. 18e-02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 LYPDAK 5
    11111
Db 6 LYPDAK 10

RESULT 10
ABG24830
ID ABG24830 standard; protein; 212 AA.
XX AC
XX ABG24830;
XX AC
XX 18-FEB-2002 (first entry)
XX DT
XX DE
XX DE Novel human diagnostic protein #24821.
XX KW
XX Human; chromosome mapping; gene mapping; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS
XX Homo sapiens.
XX PN
XX WO200175067-A2.
XX PD
XX 11-OCT-2001.
XX PF
XX 30-MAR-2001; 2001WO-US08631.
XX PR
XX 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX PA
XX Drmanac RT, Liu C, Tang YT;
XX PI

```

XX WPI: 2001-639362/73.  
DR N-PSDB: AAS89017.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PS biodiversity  
XX  
PS Claim 20; SEQ ID No 55189; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG0010-ABG3037 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
SQ Sequence 212 AA:

Query Match 100.0%; Score 26; DB 22; Length 212;  
Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 LYDAR 5  
| | | | |  
DB 37 LYDAR 41

RESULT 11  
ABBA49942  
ID ABBA49942 standard; Protein; 253 AA.  
XX  
AC ABBA49942;  
XX  
DT 05-FEB-2002 (first entry)  
XX  
DE Listeria monocytogenes protein #2646.  
XX  
KM Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
KM vitamin B12; bacterial infection; disease.  
XX  
OS Listeria monocytogenes.  
XX  
PN WO200177335-A2.  
XX  
PD 18-OCT-2001.  
XX  
PF 11-APR-2001; 2001WO-FR01118.  
XX  
PR 11-APR-2000; 2000FR-0004629.  
XX  
PA (INSP) INST PASTEUR;  
XX  
PI Buchrieser C, Frangoul L, Couve E, Rusniok C, Fshih H, Dehoux P;  
PI Dusserget O, Chetouani F, Nedjati H, Glaser P, Kunst F, Cossart P;  
PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;  
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;

PI Chakraborty T, Dommann E, Hain T, Berche P, Charbit A, Durant L;  
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;  
PI Madueno E, De Pablos B, Wehlend J, Kaerst U, Entian K, Hauf J;  
PI Rose M, Voss H;  
XX  
DR WPI: 2002-010914/01.  
XX  
PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment  
PT and prevention of Listeria and related bacterial infections, and  
PT related polypeptides  
XX  
PS Claim 6; SEQ ID No 2647; 192pp; French.  
XX  
CC The present invention relates to the genome sequence of Listeria  
CC monocytogenes EGD-e (see ABX03041). The genome sequence and fragments of  
CC it are useful for selecting probes and primers for detecting genes in L.  
CC monocytogenes and related organisms, and for studying genetic  
CC polymorphisms and other genomes. The present sequence is a protein  
CC encoded by the genome sequence of the present invention. Proteins  
CC expressed from the genome sequence are useful for raising specific  
CC antibodies, identification of L. monocytogenes and related organisms, and  
CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin  
CC B12. The genome sequence and proteins encoded by it are also useful for  
CC selecting compounds that regulate gene expression and cell replication  
CC and modulate L. monocytogenes-related diseases. In addition, the genome  
CC sequence and proteins encoded by it are useful in pharmaceutical and  
CC vaccine compositions for the treatment or prevention of infections by L.  
CC monocytogenes and related organisms.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
SQ Sequence 253 AA:

Query Match 100.0%; Score 26; DB 23; Length 253;  
Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 LYDAR 5  
| | | | |  
DB 231 LYDAR 235

RESULT 12  
AAG25581  
ID AAG25581 standard; Protein; 263 AA.  
XX  
AC AAG25581;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 29703.  
XX  
KM Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
KM termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999: 99US-0128714.  
PR 16-APR-1999: 99US-0129845.  
PR 19-APR-1999: 99US-0130077.  
PR 21-APR-1999: 99US-0130449.  
PR 23-APR-1999: 99US-0130510.  
PR 23-APR-1999: 99US-0130891.  
PR 28-APR-1999: 99US-0131449.  
PR 30-APR-1999: 99US-0132048.  
PR 04-APR-1999: 99US-0132407.  
PR 04-MAY-1999: 99US-0132484.  
PR 05-MAY-1999: 99US-0132485.  
PR 06-MAY-1999: 99US-0132486.  
PR 07-MAY-1999: 99US-0132487.  
PR 11-MAY-1999: 99US-0132863.  
PR 11-MAY-1999: 99US-0134256.  
PR 14-MAY-1999: 99US-0134218.  
PR 14-MAY-1999: 99US-0134219.  
PR 14-MAY-1999: 99US-0134221.  
PR 14-MAY-1999: 99US-0134370.  
PR 18-MAY-1999: 99US-0134768.  
PR 19-MAY-1999: 99US-0134941.  
PR 20-MAY-1999: 99US-0135124.  
PR 21-MAY-1999: 99US-0135353.  
PR 24-MAY-1999: 99US-0135629.  
PR 25-MAY-1999: 99US-0136021.  
PR 27-MAY-1999: 99US-0136392.  
PR 28-MAY-1999: 99US-0136782.  
PR 01-JUN-1999: 99US-0137222.  
PR 03-JUN-1999: 99US-0137528.  
PR 04-JUN-1999: 99US-0137502.  
PR 07-JUN-1999: 99US-0137724.  
PR 08-JUN-1999: 99US-0138094.  
PR 10-JUN-1999: 99US-0138540.  
PR 10-JUN-1999: 99US-0138847.  
PR 14-JUN-1999: 99US-0139119.  
PR 16-JUN-1999: 99US-0139452.  
PR 16-JUN-1999: 99US-0139453.  
PR 17-JUN-1999: 99US-0139492.  
PR 18-JUN-1999: 99US-0139454.  
PR 18-JUN-1999: 99US-0139455.  
PR 18-JUN-1999: 99US-0139456.  
PR 18-JUN-1999: 99US-0139457.  
PR 18-JUN-1999: 99US-0139458.  
PR 18-JUN-1999: 99US-0139459.  
PR 18-JUN-1999: 99US-0139460.  
PR 18-JUN-1999: 99US-0139461.  
PR 18-JUN-1999: 99US-0139462.  
PR 18-JUN-1999: 99US-0139463.  
PR 18-JUN-1999: 99US-0139750.  
PR 18-JUN-1999: 99US-0139763.  
PR 21-JUN-1999: 99US-0139817.  
PR 22-JUN-1999: 99US-0139899.  
PR 23-JUN-1999: 99US-0140353.  
PR 23-JUN-1999: 99US-0140354.  
PR 24-JUN-1999: 99US-0140695.  
PR 28-JUN-1999: 99US-0140823.  
PR 29-JUN-1999: 99US-0140991.  
PR 30-JUN-1999: 99US-0141287.  
PR 01-JUL-1999: 99US-0141842.  
PR 02-JUL-1999: 99US-0142154.  
PR 06-JUL-1999: 99US-0142055.  
PR 06-JUL-1999: 99US-0142390.  
PR 08-JUL-1999: 99US-0142803.  
PR 09-JUL-1999: 99US-0142820.  
PR 12-JUL-1999: 99US-0142977.  
PR 13-JUL-1999: 99US-0143542.  
PR 14-JUL-1999: 99US-0143624.  
PR 15-JUL-1999: 99US-0144005.  
PR 16-JUL-1999: 99US-0144085.  
PR 16-JUL-1999: 99US-0144086.  
PR 19-JUL-1999: 99US-0144325.  
PR 19-JUL-1999: 99US-0144331.  
PR 19-JUL-1999: 99US-0144332.  
  
PR 19-JUL-1999: 99US-0144333.  
PR 19-JUL-1999: 99US-0144334.  
PR 19-JUL-1999: 99US-0144335.  
PR 20-JUL-1999: 99US-0144332.  
PR 20-JUL-1999: 99US-0144632.  
PR 20-JUL-1999: 99US-0144684.  
PR 21-JUL-1999: 99US-0144814.  
PR 21-JUL-1999: 99US-0145086.  
PR 21-JUL-1999: 99US-0145088.  
PR 22-JUL-1999: 99US-0145085.  
PR 22-JUL-1999: 99US-0145087.  
PR 22-JUL-1999: 99US-0145089.  
PR 22-JUL-1999: 99US-0145192.  
PR 23-JUL-1999: 99US-0145145.  
PR 23-JUL-1999: 99US-0145218.  
PR 23-JUL-1999: 99US-0145224.  
PR 26-JUL-1999: 99US-0145276.  
PR 27-JUL-1999: 99US-0145913.  
PR 27-JUL-1999: 99US-0145918.  
PR 27-JUL-1999: 99US-0145919.  
PR 28-JUL-1999: 99US-0145951.  
PR 02-AUG-1999: 99US-0146386.  
PR 02-AUG-1999: 99US-0146388.  
PR 03-AUG-1999: 99US-0146389.  
PR 04-AUG-1999: 99US-0147038.  
PR 04-AUG-1999: 99US-0147204.  
PR 05-AUG-1999: 99US-0147302.  
PR 05-AUG-1999: 99US-0147192.  
PR 06-AUG-1999: 99US-0147260.  
PR 06-AUG-1999: 99US-0147303.  
PR 06-AUG-1999: 99US-0147416.  
PR 09-AUG-1999: 99US-0147493.  
PR 09-AUG-1999: 99US-0147935.  
PR 10-AUG-1999: 99US-0148171.  
PR 11-AUG-1999: 99US-0148319.  
PR 12-AUG-1999: 99US-0148341.  
PR 13-AUG-1999: 99US-0148565.  
PR 13-AUG-1999: 99US-0148684.  
PR 16-AUG-1999: 99US-0149368.  
PR 17-AUG-1999: 99US-0149175.  
PR 18-AUG-1999: 99US-0149426.  
PR 20-AUG-1999: 99US-0149722.  
PR 20-AUG-1999: 99US-0149723.  
PR 20-AUG-1999: 99US-0149929.  
PR 23-AUG-1999: 99US-0149902.  
PR 23-AUG-1999: 99US-0149930.  
PR 25-AUG-1999: 99US-0150566.  
PR 26-AUG-1999: 99US-0150884.  
PR 27-AUG-1999: 99US-0151065.  
PR 27-AUG-1999: 99US-0151066.  
PR 27-AUG-1999: 99US-0151080.  
PR 30-AUG-1999: 99US-0151303.  
PR 31-AUG-1999: 99US-0151438.  
PR 01-SEP-1999: 99US-0151930.  
PR 07-SEP-1999: 99US-0152363.  
PR 10-SEP-1999: 99US-0153070.  
PR 13-SEP-1999: 99US-0153758.  
PR 15-SEP-1999: 99US-0154018.  
PR 16-SEP-1999: 99US-0154039.  
PR 20-SEP-1999: 99US-0154779.  
PR 22-SEP-1999: 99US-0155139.  
PR 23-SEP-1999: 99US-0155486.  
PR 24-SEP-1999: 99US-0156459.  
PR 28-SEP-1999: 99US-0156458.  
PR 29-SEP-1999: 99US-0156596.  
PR 04-OCT-1999: 99US-0157117.  
PR 05-OCT-1999: 99US-0157753.  
PR 06-OCT-1999: 99US-0157865.  
PR 07-OCT-1999: 99US-0158029.  
PR 08-OCT-1999: 99US-0158232.  
PR 12-OCT-1999: 99US-0158369.  
PR 13-OCT-1999: 99US-0159293.  
PR 13-OCT-1999: 99US-0159294.

PR	13-OCT-1999;	99US-0159295;
PR	14-OCT-1999;	99US-0159329;
PR	14-OCT-1999;	99US-0159330;
PR	14-OCT-1999;	99US-0159331;
PR	14-OCT-1999;	99US-0159337;
PR	14-OCT-1999;	99US-0159638;
PR	18-OCT-1999;	99US-0159584;
PR	21-OCT-1999;	99US-0160741;
PR	21-OCT-1999;	99US-0160767;
PR	21-OCT-1999;	99US-0160768;
PR	21-OCT-1999;	99US-0160770;
PR	21-OCT-1999;	99US-0160814;
PR	21-OCT-1999;	99US-0160815;
PR	22-OCT-1999;	99US-0160980;
PR	22-OCT-1999;	99US-0160981;
PR	22-OCT-1999;	99US-0160989;
PR	25-OCT-1999;	99US-0161404;
PR	25-OCT-1999;	99US-0161405;
PR	25-OCT-1999;	99US-0161406;
PR	26-OCT-1999;	99US-0161359;
PR	26-OCT-1999;	99US-0161360;
PR	26-OCT-1999;	99US-0161361;
PR	28-OCT-1999;	99US-0161920;
PR	28-OCT-1999;	99US-0161921;
PR	28-OCT-1999;	99US-0161993;
PR	29-OCT-1999;	99US-0162142;

PR	30-APR-1999	9905-01332048
PR	30-APR-1999	9905-01332407
PR	04-MAY-1999	9905-01332484
PR	05-MAY-1999	9905-01332485
PR	06-MAY-1999	9905-01332486
PR	06-MAY-1999	9905-01332487
PR	07-MAY-1999	9905-01332488
PR	11-MAY-1999	9905-01334256
PR	14-MAY-1999	9905-01334218
PR	14-MAY-1999	9905-01334219
PR	14-MAY-1999	9905-01334221
PR	14-MAY-1999	9905-01334220
PR	18-MAY-1999	9905-01334768
PR	19-MAY-1999	9905-01334941
PR	20-MAY-1999	9905-01351524
PR	21-MAY-1999	9905-01351523
PR	24-MAY-1999	9905-01355529
PR	25-MAY-1999	9905-01336021
PR	27-MAY-1999	9905-01336392
PR	28-MAY-1999	9905-01336782
PR	01-JUN-1999	9905-01337222
PR	03-JUN-1999	9905-01337528
PR	04-JUN-1999	9905-01337502
PR	07-JUN-1999	9905-01337724
PR	08-JUN-1999	9905-01380094
PR	10-JUN-1999	9905-01338547
PR	10-JUN-1999	9905-01338540
PR	14-JUN-1999	9905-01339119
PR	16-JUN-1999	9905-01339452
PR	16-JUN-1999	9905-01339453
PR	18-JUN-1999	9905-01339457
PR	18-JUN-1999	9905-01339458
PR	18-JUN-1999	9905-01339459
PR	18-JUN-1999	9905-01339460
PR	18-JUN-1999	9905-01339461
PR	18-JUN-1999	9905-01339455
PR	18-JUN-1999	9905-01339456
PR	18-JUN-1999	9905-01339462
PR	18-JUN-1999	9905-01339463
PR	18-JUN-1999	9905-01339750
PR	18-JUN-1999	9905-01339763
PR	21-JUN-1999	9905-01339817
PR	22-JUN-1999	9905-01339869
PR	23-JUN-1999	9905-01400353
PR	23-JUN-1999	9905-01400354
PR	24-JUN-1999	9905-01400625
PR	28-JUN-1999	9905-01400623
PR	29-JUN-1999	9905-01400591
PR	30-JUN-1999	9905-01411887
PR	01-JUL-1999	9905-01418242
PR	01-JUL-1999	9905-01418254
PR	02-JUL-1999	9905-01420255
PR	06-JUL-1999	9905-01423290
PR	08-JUL-1999	9905-01423803
PR	09-JUL-1999	9905-01428270
PR	12-JUL-1999	9905-01429227
PR	13-JUL-1999	9905-01435342
PR	14-JUL-1999	9905-01435324
PR	15-JUL-1999	9905-01440005
PR	16-JUL-1999	9905-01440086
PR	16-JUL-1999	9905-01440086
PR	19-JUL-1999	9905-01444331
PR	19-JUL-1999	9905-01443332
PR	19-JUL-1999	9905-01443332
PR	19-JUL-1999	9905-01443334
PR	19-JUL-1999	9905-01443335
PR	20-JUL-1999	9905-01443525
PR	20-JUL-1999	9905-01446332
PR	20-JUL-1999	9905-01446384
PR	21-JUL-1999	9905-01446184

PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151338.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.

PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 100.0%; Score 26; DB 21; Length 263;  
Best Local Similarity 100.0%; Pred. No. 3.2e+02; Mismatches 0;  
Matches 5; Conservative 0; Indels 0; Gaps 0;

Oy 1 LYDAK 5  
Db 114 LYDAK 118

RESULT 14  
AAC25580  
ID AAC25580 standard; Protein; 297 AA.  
XX  
AC AAG25580;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 29702.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
hybridisation assay; genetic mapping; gene expression control; promoter;  
termination sequence.  
OS Arabidopsis thaliana.  
XX  
PN EPI033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 9905-0134256.  
PR 14-MAY-1999; 9905-0134218.  
PR 14-MAY-1999; 9905-0134219.  
PR 14-MAY-1999; 9905-0134221.  
PR 14-MAY-1999; 9905-0134320.  
PR 14-MAY-1999; 9905-0134370.  
PR 18-MAY-1999; 9905-0134768.  
PR 19-MAY-1999; 9905-0134941.  
PR 20-MAY-1999; 9905-0135124.  
PR 21-MAY-1999; 9905-0135353.  
PR 24-MAY-1999; 9905-0135629.  
PR 25-MAY-1999; 9905-0136021.  
PR 27-MAY-1999; 9905-0136392.  
PR 28-MAY-1999; 9905-0136782.  
PR 01-JUN-1999; 9905-0137222.  
PR 03-JUN-1999; 9905-0137528.  
PR 04-JUN-1999; 9905-0137502.  
PR 07-JUN-1999; 9905-0137724.  
PR 08-JUN-1999; 9905-0138094.  
PR 10-JUN-1999; 9905-0138540.  
PR 10-JUN-1999; 9905-0138847.  
PR 14-JUN-1999; 9905-0139119.  
PR 16-JUN-1999; 9905-0139452.  
PR 16-JUN-1999; 9905-0139453.  
PR 17-JUN-1999; 9905-0139492.  
PR 18-JUN-1999; 9905-0139454.  
PR 18-JUN-1999; 9905-0139455.  
PR 18-JUN-1999; 9905-0139456.  
PR 18-JUN-1999; 9905-0139457.  
PR 18-JUN-1999; 9905-0139458.  
PR 18-JUN-1999; 9905-0139459.  
PR 18-JUN-1999; 9905-0139460.  
PR 18-JUN-1999; 9905-0139461.  
PR 18-JUN-1999; 9905-0139462.  
PR 18-JUN-1999; 9905-0139463.  
PR 18-JUN-1999; 9905-0139750.  
PR 18-JUN-1999; 9905-0139763.  
PR 21-JUN-1999; 9905-0139817.  
PR 22-JUN-1999; 9905-0139899.  
PR 23-JUN-1999; 9905-0140353.  
PR 23-JUN-1999; 9905-0140354.  
PR 24-JUN-1999; 9905-0140695.  
PR 28-JUN-1999; 9905-0140823.  
PR 29-JUN-1999; 9905-0140991.  
PR 30-JUN-1999; 9905-0141287.  
PR 01-JUL-1999; 9905-0141842.  
PR 01-JUL-1999; 9905-0142154.  
PR 02-JUL-1999; 9905-0142055.  
PR 06-JUL-1999; 9905-0142390.  
PR 08-JUL-1999; 9905-0142803.  
PR 09-JUL-1999; 9905-0142920.  
PR 12-JUL-1999; 9905-0142977.  
PR 13-JUL-1999; 9905-0143542.  
PR 14-JUL-1999; 9905-0143624.  
PR 15-JUL-1999; 9905-0144005.  
PR 16-JUL-1999; 9905-0144085.  
PR 16-JUL-1999; 9905-0144086.  
PR 19-JUL-1999; 9905-0144325.  
PR 19-JUL-1999; 9905-0144331.  
PR 19-JUL-1999; 9905-0144332.  
PR 19-JUL-1999; 9905-0144333.  
PR 19-JUL-1999; 9905-0144334.  
PR 19-JUL-1999; 9905-0144335.  
PR 20-JUL-1999; 9905-0144352.  
PR 20-JUL-1999; 9905-0144632.  
PR 20-JUL-1999; 9905-0144884.  
PR 21-JUL-1999; 9905-0144814.  
PR 21-JUL-1999; 9905-0145086.  
PR 21-JUL-1999; 9905-0145088.  
PR 22-JUL-1999; 9905-0145085.  
PR 22-JUL-1999; 9905-0145087.  
PR 22-JUL-1999; 9905-0145089.  
PR 22-JUL-1999; 9905-0145192.  
PR 23-JUL-1999; 9905-0145145.  
  
PR 23-JUL-1999; 9905-0145218.  
PR 23-JUL-1999; 9905-0145224.  
PR 26-JUL-1999; 9905-0145276.  
PR 27-JUL-1999; 9905-0145913.  
PR 27-JUL-1999; 9905-0145918.  
PR 27-JUL-1999; 9905-0145919.  
PR 28-JUL-1999; 9905-0145951.  
PR 02-AUG-1999; 9905-0146386.  
PR 02-AUG-1999; 9905-0146388.  
PR 03-AUG-1999; 9905-0147038.  
PR 04-AUG-1999; 9905-0147204.  
PR 04-AUG-1999; 9905-0147302.  
PR 05-AUG-1999; 9905-0147192.  
PR 05-AUG-1999; 9905-0147260.  
PR 06-AUG-1999; 9905-0147303.  
PR 06-AUG-1999; 9905-0147416.  
PR 09-AUG-1999; 9905-0147493.  
PR 09-AUG-1999; 9905-0147935.  
PR 10-AUG-1999; 9905-0148171.  
PR 11-AUG-1999; 9905-0148319.  
PR 12-AUG-1999; 9905-0148341.  
PR 13-AUG-1999; 9905-0148565.  
PR 13-AUG-1999; 9905-0148684.  
PR 16-AUG-1999; 9905-0149368.  
PR 17-AUG-1999; 9905-0149175.  
PR 18-AUG-1999; 9905-0149426.  
PR 20-AUG-1999; 9905-0149722.  
PR 20-AUG-1999; 9905-0149723.  
PR 20-AUG-1999; 9905-0149729.  
PR 23-AUG-1999; 9905-0149902.  
PR 23-AUG-1999; 9905-0149930.  
PR 25-AUG-1999; 9905-0150566.  
PR 26-AUG-1999; 9905-0150884.  
PR 27-AUG-1999; 9905-0151065.  
PR 27-AUG-1999; 9905-0151066.  
PR 27-AUG-1999; 9905-0151080.  
PR 30-AUG-1999; 9905-0151303.  
PR 31-AUG-1999; 9905-0151438.  
PR 01-SEP-1999; 9905-0151930.  
PR 07-SEP-1999; 9905-0152363.  
PR 10-SEP-1999; 9905-0153070.  
PR 13-SEP-1999; 9905-0153758.  
PR 15-SEP-1999; 9905-0154018.  
PR 16-SEP-1999; 9905-0154039.  
PR 20-SEP-1999; 9905-0154779.  
PR 22-SEP-1999; 9905-0155139.  
PR 23-SEP-1999; 9905-0155486.  
PR 24-SEP-1999; 9905-0155659.  
PR 28-SEP-1999; 9905-0156458.  
PR 29-SEP-1999; 9905-0156596.  
PR 04-OCT-1999; 9905-0157117.  
PR 05-OCT-1999; 9905-0157753.  
PR 06-OCT-1999; 9905-0157865.  
PR 07-OCT-1999; 9905-0158029.  
PR 08-OCT-1999; 9905-0158232.  
PR 12-OCT-1999; 9905-0158369.  
PR 13-OCT-1999; 9905-0158293.  
PR 13-OCT-1999; 9905-0158294.  
PR 13-OCT-1999; 9905-0158295.  
PR 14-OCT-1999; 9905-0159329.  
PR 14-OCT-1999; 9905-0159330.  
PR 14-OCT-1999; 9905-0159331.  
PR 14-OCT-1999; 9905-0159637.  
PR 14-OCT-1999; 9905-0159638.  
PR 18-OCT-1999; 9905-0159638.  
PR 18-OCT-1999; 9905-0159584.  
PR 21-OCT-1999; 9905-0160741.  
PR 21-OCT-1999; 9905-0160767.  
PR 21-OCT-1999; 9905-0160768.  
PR 21-OCT-1999; 9905-0160770.  
PR 21-OCT-1999; 9905-0160814.  
PR 21-OCT-1999; 9905-0160815.  
PR 22-OCT-1999; 9905-0160980.

PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 100.0%; Score 26; DB 21; Length 297;  
Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5  
|||||  
Db 148 LYDAK 152

RESULT 15  
AAG37819  
ID AAG37819 standard; Protein: 297 AA.  
XX AAG37819;  
AC  
XX  
DT 18-OCT-2000 (first entry)  
XX  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 46565.  
DE  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
XX Arabidopsis thaliana.  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134318.  
PR 14-MAY-1999; 99US-0134319.  
PR 14-MAY-1999; 99US-0134321.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139453.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144844.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 26-JUL-1999; 99US-0145313.  
PR 27-JUL-1999; 99US-0145318.  
PR 27-JUL-1999; 99US-0145319.  
PR 28-JUL-1999; 99US-0145951.

PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 18-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.

PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 100.0%; Score 26; DB 21; Length 297;  
Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5  
DB 148 LYDAK 152

Search completed: July 11, 2003, 11:54:24  
Job time : 4.86902 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW protein - protein search, using SW model

Run on: July 11, 2003, 11:49:36 ; Search time 0.893971 Seconds  
(Without alignments)  
164.563 Million cell updates/sec

Title: US-10-091-442-7

Perfect score: 26

Sequence: 1 LVDK 5

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed.  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	26	100.0	5	2	US-08-472-659-7
2	26	100.0	5	2	US-08-474-661-7
3	26	100.0	5	2	US-08-611-977-7
4	26	100.0	380	2	US-08-472-659-34
5	26	100.0	380	2	US-08-474-661-34
6	26	100.0	380	2	US-08-611-977-34
7	26	100.0	630	4	US-09-300-909-20
8	24	92.3	88	4	US-08-936-165A-513
9	24	92.3	377	3	US-09-023-023-1
10	24	92.3	388	2	US-08-742-621-1
11	24	92.3	388	4	US-09-191-608-22
12	24	92.3	471	1	US-08-203-905B-2
13	24	92.3	472	1	US-08-203-905B-14
14	24	92.3	586	4	US-09-134-001C-4456
15	23	88.5	11	4	US-09-177-249-236
16	23	88.5	75	4	US-08-858-207A-460
17	23	88.5	77	4	US-08-803-346-64
18	23	88.5	155	4	US-09-129-030-6
19	23	88.5	178	4	US-09-129-030-54
20	23	88.5	185	4	US-09-387-800-2
21	23	88.5	284	2	US-08-751-233A-6
22	23	88.5	284	2	US-08-751-233A-8
23	23	88.5	308	4	US-09-323-872A-34
24	23	88.5	434	4	US-08-370-190-3
25	23	88.5	434	4	US-09-236-615-3
26	23	88.5	436	2	US-08-537-400-3
27	23	88.5	452	2	US-08-731-079A-2

28	23	88.5	547	4	US-09-178-252-11	Sequence 11, Appl
29	23	88.5	569	1	US-08-306-231-3	Sequence 3, Appl
30	23	88.5	585	2	US-08-426-125-2	Sequence 2, Appl
31	23	88.5	585	2	US-08-455-355-2	Sequence 2, Appl
32	23	88.5	876	1	US-08-717-515-4	Sequence 4, Appl
33	23	88.5	893	2	US-08-706-702-3	Sequence 3, Appl
34	23	88.5	893	3	US-08-484-661A-2	Sequence 2, Appl
35	23	88.5	893	3	US-08-706-706-3	Sequence 3, Appl
36	23	88.5	893	3	US-08-656-664-2	Sequence 2, Appl
37	23	88.5	893	4	US-09-019-160-2	Sequence 2, Appl
38	23	88.5	893	4	US-09-019-160-6	Sequence 6, Appl
39	23	88.5	893	4	US-09-019-160-7	Sequence 7, Appl
40	23	88.5	893	4	US-09-019-160-8	Sequence 8, Appl
41	23	88.5	893	4	US-09-019-160-9	Sequence 9, Appl
42	23	88.5	893	5	PCT-US96-09641-2	Sequence 2, Appl
43	23	88.5	896	2	US-08-640-389A-10	Sequence 10, Appl
44	23	88.5	896	2	US-08-640-389A-12	Sequence 12, Appl
45	23	88.5	896	4	US-08-618-957A-10	Sequence 10, Appl

#### ALIGNMENTS

RESULT 1  
US-08-472-659-7  
Sequence 7, Application US/08472659  
Patent No. 5831030  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSUROUOKA, No. 5831030uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. 5831030uhito  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAUCHI, Kozo  
APPLICANT: YAMAGUCHI, No. 5831030omi  
TITLE OF INVENTION: MECAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/472, 659  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091, 028  
FILING DATE: 14-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-248  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:

LENGTH: 5 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-472-659-7

Query Match 100.0%; Score 26; DB 2; Length 5;  
Best Local Similarity 100.0%; Pred. No. 1.9e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5  
11111  
DB 1 LYDAK 5

RESULT 2  
US-08-474-661-7  
Sequence 7, Application US/08474661  
Patent No. 5874253

GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSURUOKA, No. 5874253uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. 5874253uhiro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAGUCHI, Kozo  
APPLICANT: YAMAGUCHI, No. 5874253omi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/474, 661  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: REA, TERESA STANEK  
REGISTRATION NUMBER: 30,427  
REFERENCE/DOCKET NUMBER: 001560-204  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-6620  
INFORMATION FOR SEQ. ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-474-661-7

Query Match 100.0%; Score 26; DB 2; Length 5;

Best Local Similarity 100.0%; Pred. No. 1.9e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5  
11111  
DB 1 LYDAK 5

RESULT 3  
US-08-611-977-7  
Sequence 7, Application US/08611977  
Patent No. 5972886

GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSURUOKA, No. 5972886uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. 5972886uhiro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAGUCHI, Kozo  
APPLICANT: YAMAGUCHI, No. 5972886omi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/611, 977  
FILING DATE: 06-MAR-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: MCGOWAN, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-204  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ. ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-611-977-7

Query Match 100.0%; Score 26; DB 2; Length 5;  
Best Local Similarity 100.0%; Pred. No. 1.9e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5  
11111  
DB 1 LYDAK 5

RESULT 4  
US-08-472-659-34  
Sequence 34, Application US/08472659  
Patent No. 5831030  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSUROOKA, No. 5831030uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. 5831030uhiro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, No. 5831030omi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
City: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/472,659  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-248  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-6620  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-472-659-34

Query Match 100.0%; Score 26; DB 2; Length 380;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5  
DB 121 LYDAK 125

RESULT 5  
US-08-474-661-34  
Sequence 34, Application US/08474661  
Patent No. 5874253  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki

APPLICANT: TSUROOKA, No. 5874253uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. 5874253uhiro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, No. 5874253omi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
City: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/474,661  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: REA, TERESA STANEK  
REGISTRATION NUMBER: 30,427  
REFERENCE/DOCKET NUMBER: 001560-204  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-6620  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-474-661-34

Query Match 100.0%; Score 26; DB 2; Length 380;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5  
DB 121 LYDAK 125

RESULT 6  
US-08-611-977-34  
Sequence 34, Application US/08611977  
Patent No. 5972886  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSUROOKA, No. 5972886uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. 5972886uhiro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, No. 5972886omi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34

;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Burns, Doane, Swecker & Mathis  
;; STREET: P.O. Box 1404  
;; CITY: Alexandria  
;; STATE: Virginia  
;; COUNTRY: United States  
;; ZIP: 22113-1404  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patent Release #1.0, Version #1.30  
;;  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/611,977  
;; FILING DATE: 06-MAR-1996  
;; CLASSIFICATION: 435  
;;  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/091,028  
;; FILING DATE: 14-JUL-1993  
;;  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: JP 4-212305  
;; FILING DATE: 17-JUL-1992  
;;  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: JP 6-067339  
;; FILING DATE: 04-MAR-1993  
;;  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: MCGOWAN, Malcolm K.  
;; REGISTRATION NUMBER: 39,300  
;; REFERENCE/DOCKET NUMBER: 001560-204  
;;  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (703) 836-6620  
;; TELEFAX: (703) 836-2021  
;;  
;; INFORMATION FOR SEQ ID NO: 34:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 380 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;;  
;; US-08-611-977-34

Query Match 100.0%; Score 26; DB 2; Length 380;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5  
DB 121 LYDAK 125

RESULT 7  
US-09-300-909-20  
; Sequence 20, Application US/09300909  
; Patent No. 6306580  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: PREPARATION OF HUMAN PAPILLOMAVIRUS E1 HAVING  
; NUMBER OF SEQUENCES: 27  
; NUMBER OF SEQUENCES: 27  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/300,909  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/083,942  
; FILING DATE: 01-MAY-1998  
; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 630 amino acids  
; TYPE: amino acid

;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;;  
;; US-09-300-909-20

Query Match 100.0%; Score 26; DB 4; Length 630;  
Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5  
DB 493 LYDAK 497

RESULT 8  
US-08-936-165A-513  
; Sequence 513, Application US/08936165A  
; Patent No. 6348582  
; GENERAL INFORMATION:  
; APPLICANT: Black, Michael  
; APPLICANT: Burnham, Martin  
; APPLICANT: Hodgson, John  
; APPLICANT: Knowles, David  
; APPLICANT: Lonetto, Michael  
; APPLICANT: Nicholas, Richard  
; APPLICANT: Pratt, Julie  
; APPLICANT: Reichard, Richard  
; APPLICANT: Rosenberg, Martin  
; APPLICANT: Ward, Judith  
; TITLE OF INVENTION: No. 6348582e1 Prokaryotic Polynucleotides,  
; NUMBER OF SEQUENCES: 534  
; NUMBER OF SEQUENCES: 534  
; INFORMATION FOR SEQ ID NO: 513:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 88 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-936-165A-513

Query Match 92.3%; Score 24; DB 4; Length 88;  
Best Local Similarity 80.0%; Pred. No. 1.2e+02;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5

Db :||||  
48 MYDAK 52

## RESULT 9

US-09-023-023-2

Sequence 2, Application US/09023023

Patent No. 6121018

GENERAL INFORMATION:

APPLICANT: Kristine Kay Kikly

TITLE OF INVENTION: Interleukin-1 Beta Converting Enzyme Like Apoptosis Pr

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: SmithKline Beecham Corporation

STREET: 709 Swedeland Road

CITY: King of Prussia

STATE: PA

COUNTRY: USA

ZIP: 19406

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/023,023

FILING DATE: 12-FEB-1998

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/042,030

FILING DATE: March 27, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Han, William T

REGISTRATION NUMBER: 34,344

REFERENCE/DOCKET NUMBER: GH50013

TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-270-5219

TELEFAX: 610-270-5090

TELEX:

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 377 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

## US-09-023-023-2

## Query Match

Best Local Similarity 92.3%; Score 24; DB 3; Length 377;

Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

## Db 1 LYDAK 5

45 IYDAK 49

## RESULT 10

US-08-742-621-1

Sequence 1, Application US/08742621

Patent No. 5856129

GENERAL INFORMATION:

APPLICANT: HILLMAN, JENNIFER L.

TITLE OF INVENTION: NOVEL HUMAN PURINOCEPTOR

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: US

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ Version 1.5

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/742,621

FILING DATE: Filed Herewith

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0147 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 388 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

IMMEDIATE SOURCE:

LIBRARY:

CLONE: consensus

## US-08-742-621-1

## Query Match

Best Local Similarity 92.3%; Score 24; DB 2; Length 388;

Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

## Db 1 LYDAK 5

218 IYDAK 222

## RESULT 11

US-09-191-608-22

Sequence 22, Application US/09191608

Patent No. 6242216

GENERAL INFORMATION:

APPLICANT: Lynch, Kevin J.

APPLICANT: Burgard, Edward C.

APPLICANT: Metzger, Randy E.

APPLICANT: Niforatos, Wende

APPLICANT: Touma, Edward B.

APPLICANT: Van Biesen, T.

TITLE OF INVENTION: Nucleic Acids Encoding a Functional

TITLE OF INVENTION: Human Purinoceptor P2X2 and P2X4 and Methods of Production

FILE REFERENCE: 6394.US.P1

CURRENT APPLICATION NUMBER: US/09/191,608

CURRENT FILING DATE: 1998-11-13

NUMBER OF SEQ ID NOS: 26

SOFTWARE: FASTSEQ for Windows Version 3.0

SEQ ID NO 22

LENGTH: 388

TYPE: PRT

ORGANISM: Homo sapiens

## US-09-191-608-22

## Query Match

Best Local Similarity 92.3%; Score 24; DB 4; Length 388;

Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

## Db 1 LYDAK 5

218 IYDAK 222

## RESULT 12

US-08-203-905B-2  
Sequence 2, Application US/08203905B  
Patent No. 5646249  
GENERAL INFORMATION:  
APPLICANT: KAYE, FEDERIC J.  
APPLICANT: OTTERSON, GREGORY A.  
TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF A  
TITLE OF INVENTION: NOVEL CHAPERONE PROTEIN  
NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: KNOBBE, MARTENS, OLSON & BEAR  
STREET: 620 NEWPORT CENTER DRIVE, SIXTEENTH FLOOR  
CITY: NEWPORT BEACH  
STATE: CA  
COUNTRY: USA  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/203,905B  
FILING DATE: February 28, 1994  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: KIRKPATRICK, ANITA M.  
REGISTRATION NUMBER: 32,617  
REFERENCE/DOCKET NUMBER: NIH089,001A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-235-8550  
TELEFAX: 619-235-0176  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 471 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-203-905B-2

Query Match  
Best Local Similarity 92.3%; Score 24; DB 1; Length 471;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5  
:||||  
Db 96 IYDAK 100

RESULT 13  
US-08-203-905B-14  
Sequence 14, Application US/08203905B  
Patent No. 5646249  
GENERAL INFORMATION:  
APPLICANT: KAYE, FEDERIC J.  
APPLICANT: OTTERSON, GREGORY A.  
TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF A  
TITLE OF INVENTION: NOVEL CHAPERONE PROTEIN  
NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: KNOBBE, MARTENS, OLSON & BEAR  
STREET: 620 NEWPORT CENTER DRIVE, SIXTEENTH FLOOR  
CITY: NEWPORT BEACH  
STATE: CA  
COUNTRY: USA  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/203,905B

FILING DATE: February 28, 1994  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: KIRKPATRICK, ANITA M.  
REGISTRATION NUMBER: 32,617  
REFERENCE/DOCKET NUMBER: NIH089,001A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-235-8550  
TELEFAX: 619-235-0176  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 472 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-203-905B-14

Query Match  
Best Local Similarity 92.3%; Score 24; DB 1; Length 472;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5  
:||||  
Db 97 IYDAK 101

RESULT 14  
US-09-134-001C-4456  
Sequence 4456, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC  
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 4456  
LENGTH: 586  
TYPE: PRN  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4456

Query Match  
Best Local Similarity 92.3%; Score 24; DB 4; Length 586;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5  
:||||  
Db 94 IYDAK 98

RESULT 15  
US-09-177-249-236  
Sequence 236, Application US/09177249  
Patent No. 6229064  
GENERAL INFORMATION:  
APPLICANT: Fischer, Robert L.  
APPLICANT: Ohad, Nir  
APPLICANT: Kiyosue, Tomohiro  
APPLICANT: Yadegari, Ramlin  
APPLICANT: Margossian, Linda  
APPLICANT: Harada, John  
APPLICANT: Goldberg, Robert B.  
TITLE OF INVENTION: The Regents of the University of California  
TITLE OF INVENTION: Nucleic Acids That Control Seed and Fruit  
FILE REFERENCE: 023070-08612005

; CURRENT APPLICATION NUMBER: US/09/177,249  
 ; CURRENT FILING DATE: 1998-10-22  
 ; EARLIER APPLICATION NUMBER: US 09/071,838  
 ; EARLIER FILING DATE: 1998-05-01  
 ; NUMBER OF SEQ ID NOS: 324  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 236  
 ; LENGTH: 11  
 ; TYPE: PRT  
 ; ORGANISM: Arabidopsis sp.  
 ; US-09-177-249-236

Query Match 88.5%; Score 23; DB 4; Length 11;  
 Best Local Similarity 80.0%; Pred. No. 22;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 LYDAK 5  
 |||:|  
 Db 1 LYDSK 5

Search completed: July 11, 2003, 12:02:10  
 Job time : 1.89397 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:59:17 ; Search time 1.4657 Seconds  
(without alignments)  
397.148 Million cell updates/sec

Title: US-10-091-442-7

Perfect score: 26

Sequence: 1 LYDAK 5

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 445758 seqs, 116419773 residues

Total number of hits satisfying chosen parameters: 445758

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications-AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/PCT05\_PUBCOMB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	26	100.0	5	9	US-10-091-442-7
2	26	100.0	5	10	US-09-140-719-7
3	26	100.0	380	9	US-10-091-442-34
4	26	100.0	380	10	US-09-140-719-34
5	26	100.0	413	9	US-09-738-626-4157
6	26	100.0	430	9	US-10-156-761-14800
7	26	100.0	464	9	US-10-036-9598-9
8	26	100.0	491	10	US-09-794-098-2
9	26	100.0	567	9	US-10-156-761-7781
10	26	100.0	707	9	US-10-156-761-7899
11	26	100.0	712	9	US-10-156-761-15026
12	26	100.0	745	10	US-09-815-242-12209
13	26	100.0	1014	9	US-09-738-626-4656
14	26	100.0	1043	9	US-10-242-056-61
15	24	92.3	88	10	US-09-939-980-513
16	24	92.3	91	9	US-10-156-733-10
17	24	92.3	140	12	US-10-010-901-27
18	24	92.3	347	10	US-09-816-028A-17
19	24	92.3	388	10	US-09-833-082-2

20	24	92.3	447	9	US-10-059-585-51	Sequence 51, Appl
21	24	92.3	471	9	US-10-171-311-212	Sequence 212, App
22	24	92.3	579	10	US-09-815-242-5821	Sequence 5821, Ap
23	24	92.3	579	10	US-09-815-242-12943	Sequence 12943, A
24	24	92.3	645	9	US-10-156-761-12752	Sequence 12752, A
25	24	92.3	667	9	US-09-563-728A-30	Sequence 30, Appl
26	24	92.3	967	10	US-09-817-913-7	Sequence 7, Appl1
27	24	92.3	967	10	US-09-817-538-7	Sequence 7, Appl1
28	24	92.3	1084	9	US-10-072-094-7	Sequence 7, Appl1
29	24	92.3	1084	9	US-10-173-539-12	Sequence 12, Appl
30	23	88.5	11	9	US-10-213-512-236	Sequence 236, App
31	23	88.5	11	10	US-09-071-838-236	Sequence 236, App
32	23	88.5	185	9	US-10-234-329-15	Sequence 15, Appl
33	23	88.5	434	9	US-10-112-039-3	Sequence 3, Appl1
34	23	88.5	442	10	US-09-815-242-13237	Sequence 13237, A
35	23	88.5	442	10	US-09-815-242-13702	Sequence 13702, A
36	23	88.5	444	9	US-10-108-605-187	Sequence 187, App
37	23	88.5	448	10	US-09-815-242-13272	Sequence 13272, A
38	23	88.5	454	10	US-09-815-242-10700	Sequence 10700, A
39	23	88.5	498	9	US-10-083-357-1290	Sequence 1290, Ap
40	23	88.5	518	10	US-09-816-248-17	Sequence 17, Appl
41	23	88.5	547	10	US-09-826-660-11	Sequence 11, Appl
42	23	88.5	590	10	US-09-815-242-5099	Sequence 5099, Ap
43	23	88.5	616	9	US-10-126-279-32	Sequence 32, Appl
44	23	88.5	617	9	US-10-126-279-31	Sequence 31, Appl
45	23	88.5	634	10	US-09-816-248-13	Sequence 13, Appl

#### ALIGNMENTS

RESULT 1  
US-10-091-442-7  
Sequence 7, Application US/10091442  
Patent No. US20020164711A1  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
INASA, Fuyuki  
TSURUOKA, No. US20020164711A1uo  
NAKAZATO, Hiroshi  
MURA, Kenju  
ISHIDA, No. US20020164711A1uhfro  
KURIHARA, Tatsuya  
YAMAIICHI, Kozo  
YAMAGUCHI, No. US20020164711A1om1  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/091,442  
FILING DATE: 07-Mar-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
APPLICATION NUMBER: JP 6-067339

FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: MCGOWAN, MALCOLM K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021.  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 7:  
US-10-091-442-7

Query Match 100.0%; Score 26; DB 9; Length 5;  
Best Local Similarity 100.0%; Pred. No. 4e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5  
11111  
Db 1 LYDAK 5

## RESULT 2

US-09-140-719-7  
Sequence 7, Application US/09140719  
Patent No. US20010026931A1  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSUROOKA, No. US20010026931A1uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. US20010026931A1uhiro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAGUCHI, Kozo  
APPLICANT: YAMAGUCHI, No. US20010026931A1oml  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: MCGOWAN, MALCOLM K.

REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-140-719-7

Query Match 100.0%; Score 26; DB 10; Length 5;  
Best Local Similarity 100.0%; Pred. No. 4e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5  
11111  
Db 1 LYDAK 5

## RESULT 3

US-10-091-442-34  
Sequence 34, Application US/10091442  
Patent No. US20020164711A1  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSUROOKA, No. US20020164711A1uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. US20020164711A1uhiro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAGUCHI, Kozo  
APPLICANT: YAMAGUCHI, No. US20020164711A1oml  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/091,442  
FILING DATE: 07-Mar-2002  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: MCGOWAN, MALCOLM K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021

INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 34:  
US-10-091-442-34

Query Match 100.0%; Score 26; DB 9; Length 380;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5  
DB 121 LYDAK 125

RESULT 4  
US-09-140-719-34  
Sequence 34, Application US/09140719  
Patent No. US20010026931A1  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: INASA, Fuyuki  
APPLICANT: TSUBOUOKA, No. US20010026931A1uo  
APPLICANT: MAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. US20010026931A1uh1ro  
APPLICANT: KURIHARA, Tetsuya  
APPLICANT: YAMAUCHI, Kozo  
APPLICANT: YAMAGUCHI, No. US20010026931A1om1  
TITLE OF INVENTION: MECAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
City: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
APPLICATION NUMBER: JP 6-067339  
APPLICATION DATA:  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid

TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-140-719-34

Query Match 100.0%; Score 26; DB 10; Length 380;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5  
DB 121 LYDAK 125

RESULT 5  
US-09-738-626-4157  
Sequence 4157, Application US/09738626  
Publication No. US20020197605A1  
GENERAL INFORMATION:  
APPLICANT: NAKAGAWA, SATOSHI  
APPLICANT: MIZOGUCHI, HIROSHI  
APPLICANT: ANDO, SEIKO  
APPLICANT: HAYASHI, MIKIRO  
APPLICANT: OCHIAI, KEIKO  
APPLICANT: YOKOI, HARUHIKO  
APPLICANT: TATEISHI, NAOKO  
APPLICANT: SENOH, AKIHITO  
APPLICANT: IKEDA, MASATO  
APPLICANT: OZAKI, AKIO  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-125  
CURRENT APPLICATION NUMBER: US/09/738,626  
CURRENT FILING DATE: 2000-12-18  
PRIOR APPLICATION NUMBER: JP 99/377484  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: JP 00/159162  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: JP 00/280988  
PRIOR FILING DATE: 2000-08-03  
NUMBER OF SEQ ID NOS: 7059  
SOFTWARE: PatentIn ver. 3.0  
SEQ ID NO 413  
LENGTH: 413  
TYPE: PRT  
ORGANISM: Corynebacterium glutamicum  
US-09-738-626-4157

Query Match 100.0%; Score 26; DB 9; Length 413;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5  
DB 340 LYDAK 344

RESULT 6  
US-10-156-761-14800  
Sequence 14800, Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30

;; PRIOR APPLICATION NUMBER: JP 2001-272697  
;; PRIOR FILING DATE: 2001-08-02  
;; NUMBER OF SEQ ID NOS: 15109  
;; SEQ ID NO 14800  
;; LENGTH: 430  
;; TYPE: PRN  
;; ORGANISM: Streptomyces avermitilis  
US-10-156-761-14800

Query Match  
Best Local Similarity 100.0%; Score 26; DB 9; Length 430;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5  
DB 242 LYDAK 246

RESULT 7  
US-10-036-959B-9  
;; Sequence 9, Application US/10036959B  
;; Publication No. US20030119098A1  
;; GENERAL INFORMATION:  
;; APPLICANT: E.I. du Pont de Nemours & Company  
;; APPLICANT: Hallahan, David L.  
;; APPLICANT: Keiper-Hrynko, Natalie  
;; TITLE OF INVENTION: Genes involved in the Biosynthesis of Isopentenyl Diphosphate in  
;; TITLE OF INVENTION: brasiliensis latex  
;; FILE REFERENCE: CI-1792  
;; CURRENT APPLICATION NUMBER: US/10/036,959B  
;; CURRENT FILING DATE: 2002-05-10  
;; PRIOR APPLICATION NUMBER: 60/307,637  
;; PRIOR FILING DATE: 2001-07-25  
;; NUMBER OF SEQ ID NOS: 16  
;; SOFTWARE: Microsoft Office 97  
;; SEQ ID NO 9  
;; LENGTH: 464  
;; TYPE: PRN  
;; ORGANISM: Hevea brasiliensis  
US-10-036-959B-9

Query Match  
Best Local Similarity 100.0%; Score 26; DB 9; Length 464;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5  
DB 309 LYDAK 313

RESULT 8  
US-09-794-098-2  
;; Sequence 2, Application US/09794098  
;; Patent No. US20020012668A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Universiti Sains Malaysia  
;; TITLE OF INVENTION: DNA SEQUENCE ENCODING THE SPECIFIC AND ANTIGENIC OUTER MEMBRANE  
;; FILE REFERENCE: 4463-005  
;; CURRENT APPLICATION NUMBER: US/09/794,098  
;; CURRENT FILING DATE: 2001-02-28  
;; PRIOR APPLICATION NUMBER: P/ 2000 0765  
;; PRIOR FILING DATE: 2000-02-28  
;; NUMBER OF SEQ ID NOS: 2  
;; SOFTWARE: Patentin version 3.0  
;; SEQ ID NO 2  
;; LENGTH: 491  
;; TYPE: PRN  
;; ORGANISM: Salmonella typhi  
US-09-794-098-2

Query Match  
Best Local Similarity 100.0%; Score 26; DB 10; Length 491;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 LYDAK 5  
DB 400 LYDAK 404

RESULT 9  
US-10-156-761-7781  
;; Sequence 7781, Application US/10156761  
;; Publication No. US20030119018A1  
;; GENERAL INFORMATION:  
;; APPLICANT: OMURA, SATOSHI  
;; APPLICANT: IKEDA, HARUO  
;; APPLICANT: ISHIKAWA, JUN  
;; APPLICANT: HORIKAWA, HIROSHI  
;; APPLICANT: SHIBA, TADAYOSHI  
;; APPLICANT: SAKAKI, YOSHIYUKI  
;; APPLICANT: HATTORI, MASAHIRA  
;; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
;; FILE REFERENCE: 249-262  
;; CURRENT APPLICATION NUMBER: US/10/156,761  
;; CURRENT FILING DATE: 2002-05-29  
;; PRIOR APPLICATION NUMBER: JP 2001-204089  
;; PRIOR FILING DATE: 2001-05-30  
;; PRIOR APPLICATION NUMBER: JP 2001-272697  
;; PRIOR FILING DATE: 2001-08-02  
;; NUMBER OF SEQ ID NOS: 15109  
;; SEQ ID NO 7781  
;; LENGTH: 567  
;; TYPE: PRN  
;; ORGANISM: Streptomyces avermitilis  
US-10-156-761-7781

Query Match  
Best Local Similarity 100.0%; Score 26; DB 9; Length 567;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5  
DB 447 LYDAK 451

RESULT 10  
US-10-156-761-7899  
;; Sequence 7899, Application US/10156761  
;; Publication No. US20030119018A1  
;; GENERAL INFORMATION:  
;; APPLICANT: OMURA, SATOSHI  
;; APPLICANT: IKEDA, HARUO  
;; APPLICANT: ISHIKAWA, JUN  
;; APPLICANT: HORIKAWA, HIROSHI  
;; APPLICANT: SHIBA, TADAYOSHI  
;; APPLICANT: SAKAKI, YOSHIYUKI  
;; APPLICANT: HATTORI, MASAHIRA  
;; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
;; FILE REFERENCE: 249-262  
;; CURRENT APPLICATION NUMBER: US/10/156,761  
;; CURRENT FILING DATE: 2002-05-29  
;; PRIOR APPLICATION NUMBER: JP 2001-204089  
;; PRIOR FILING DATE: 2001-05-30  
;; PRIOR APPLICATION NUMBER: JP 2001-272697  
;; PRIOR FILING DATE: 2001-08-02  
;; NUMBER OF SEQ ID NOS: 15109  
;; SEQ ID NO 7899  
;; LENGTH: 707  
;; TYPE: PRN  
;; ORGANISM: Streptomyces avermitilis  
US-10-156-761-7899

Query Match  
Best Local Similarity 100.0%; Score 26; DB 9; Length 707;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5  
DB 470 LYDAK 474

## RESULT 11

US-10-156-761-15026  
; Sequence 15026, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 15026  
; LENGTH: 712  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-15026

Query Match 100.0%; Score 26; DB 9; Length 712;  
Best Local Similarity 100.0%; Pred. No. 5.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5  
DB 472 LYDAK 476

## RESULT 12

US-09-815-242-12209  
; Sequence 12209, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 12209  
; LENGTH: 745  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-12209

Query Match 100.0%; Score 26; DB 10; Length 745;  
Best Local Similarity 100.0%; Pred. No. 5.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5  
DB 633 LYDAK 637

## RESULT 13

US-09-738-626-4656  
; Sequence 4656, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MITOUCUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENOH, AKIHITO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OKAI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 4656  
; LENGTH: 1014  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-4656

Query Match 100.0%; Score 26; DB 9; Length 1014;  
Best Local Similarity 100.0%; Pred. No. 7.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5  
DB 730 LYDAK 734

## RESULT 14

US-10-242-056-61  
; Sequence 61, Application US/10242056  
; Publication No. US20030113323A1  
; GENERAL INFORMATION:  
; APPLICANT: Ensign, Jerald C  
; APPLICANT: Bowen, David J  
; APPLICANT: Petrell, James  
; APPLICANT: Fellig, Raymond  
; APPLICANT: Schoonover, Sue  
; APPLICANT: Firench-Constant, Richard  
; APPLICANT: Orr, Gregory L  
; APPLICANT: Merlo, Donald J

APPLICANT: Roberts, Jean L  
APPLICANT: Rocheleau, Thomas A  
TITLE OF INVENTION: Insecticidal Protein Toxins from  
TITLE OF INVENTION: Photorhabdus  
NUMBER OF SEQUENCES: 88  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dowelanco  
STREET: 9330 Zionsville Road  
CITY: Indianapolis  
STATE: IN  
COUNTRY: US  
ZIP: 46268  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/242,056  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/063,615  
FILING DATE: 18-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/395,497  
FILING DATE: 28-FEB-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/007,255  
FILING DATE: 06-NOV-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/608,423  
FILING DATE: 28-FEB-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/705,484  
FILING DATE: 28-AUG-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/743,699  
FILING DATE: 06-NOV-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Borucki, Andrea T.  
REGISTRATION NUMBER: 33651  
REFERENCE/DOCKET NUMBER: 50301E  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-337-4846  
TELEFAX: 317-337-4847  
INFORMATION FOR SEQ ID NO: 61:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1043 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-10-242-056-61

Query Match 100.0%; Score 26; DB 9; Length 1043;  
Best Local Similarity 100.0%; Pred. NO. 7.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5  
:|||||  
DB 62 LYDAK 66

RESULT 15  
US-09-939-980-513  
Sequence 513, Application US/09939980  
Patent No. US20020082234A1  
GENERAL INFORMATION:  
APPLICANT: Black, Michael  
Burnham, Martin  
Hodgson, John  
Knowles, David  
Lonetto, Michael

Nicholas, Richard  
Priat, Julie  
Reichard, Richard  
Rosenberg, Martin  
Ward, Judith  
TITLE OF INVENTION: No. US20020082234A1 Prokaryotic Polynucleotides,  
Polypeptides and Their Uses  
NUMBER OF SEQUENCES: 534  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-0939  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/939,980  
FILING DATE: 27-Aug-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/936,165  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Gimm, Edward R.  
REGISTRATION NUMBER: 38,891  
REFERENCE/DOCKET NUMBER: P50549  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-4478  
TELEFAX: 610-270-5090  
TELEX: <Unknown>  
INFORMATION FOR SEQ ID NO: 513:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 88 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 513:  
US-09-939-980-513

Query Match 92.3%; Score 24; DB 10; Length 88;  
Best Local Similarity 80.0%; Pred. NO. 1.8e+02;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5  
:|||||  
DB 48 MYDAK 52

Search completed: July 11, 2003, 12:37:34  
Job time : 2.4657 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: July 11, 2003, 11:45:06 ; Search time 1.00832 seconds  
(without alignments)  
476.708 Million cell updates/sec

Title: US-10-091-442-7

Perfect score: 26

Sequence: 1 LYDAK 5

## Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

PIR\_73:\*  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	26	100.0	117	2 S41811	Ig kappa chain V r
2	26	100.0	178	2 H70345	conserved hypochet
3	26	100.0	197	2 B71106	hypothetical prote
4	26	100.0	199	2 T24871	hypothetical prote
5	26	100.0	253	2 AD1797	partition protein,
6	26	100.0	253	2 AF1423	partition protein,
7	26	100.0	259	2 G81427	periplasmic protei
8	26	100.0	277	2 S73522	probable lipoprote
9	26	100.0	281	2 T50351	probable methyltra
10	26	100.0	301	2 G97187	UDP-glucose 4-epim
11	26	100.0	304	2 C90453	hypothetical prote
12	26	100.0	326	2 T15194	hypothetical prote
13	26	100.0	341	2 B83298	hypothetical prote
14	26	100.0	356	2 E75266	conserved hypochet
15	26	100.0	401	2 T50679	GGDEF family prote
16	26	100.0	408	2 S57148	CP49 protein - chl
17	26	100.0	421	2 H89916	hypothetical prote
18	26	100.0	439	2 E98139	diaminopimelate de
19	26	100.0	439	2 AF3148	hypothetical prote
20	26	100.0	443	2 A27993	hypothetical prote
21	26	100.0	450	2 T50631	methyl coenzyme M
22	26	100.0	474	2 T09688	CP49ins protein -
23	26	100.0	475	2 T27811	hydroxymethylgluta
24	26	100.0	491	2 S70190	hypothetical prote
25	26	100.0	491	2 AB0890	tolC protein - Sal
26	26	100.0	512	2 B82433	outer membrane pro
27	26	100.0	526	1 C71081	aerobic glycerol-3
28	26	100.0	543	2 T50905	probable helicase
29	26	100.0	549	2 A90253	proteohlorophyllid

30	26	100.0	552	2 B90476	hypothetical prote
31	26	100.0	572	2 S52594	asparagine synthas
32	26	100.0	574	2 G69769	pyruvate oxidase h
33	26	100.0	577	2 D83473	isocitrate dehydro
34	26	100.0	605	2 G95853	probable pyruvate
35	26	100.0	630	1 W1ML35	E1 protein - human
36	26	100.0	633	2 H95932	probable adenylyl-
37	26	100.0	637	2 S36523	E1 protein - human
38	26	100.0	641	1 Z2RNO	adenylyl-sulfate k
39	26	100.0	641	2 E95320	adenylyl-sulfate k
40	26	100.0	662	2 T46228	hypothetical prote
41	26	100.0	730	2 S64998	hypothetical prote
42	26	100.0	734	2 AF2001	hypothetical prote
43	26	100.0	745	2 A89770	hypothetical prote
44	26	100.0	768	2 S76977	pled-4 protein - S
45	26	100.0	815	2 E70021	3-hydroxyacyl-CoA

## ALIGNMENTS

## RESULT 1

S41811

Ig kappa chain V region L22 - human

C:Species: Homo sapiens (man)

C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 21-Jul-2000

C:Accession: S41811

R:Huber, C.; Schaeble, K.F.; Huber, E.; Klein, R.; Melndl, A.; Thiebe, R.; Lamm, R.;

Eur. J. Immunol. 23, 2868-2875, 1993

A:Title: The V(kappa) genes of the L regions and the repertoire of V(kappa) gene sequ

A:Reference number: S41809; PMID:94039386; PMID:8223863

A:Accession: S41811

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-117 <HUB>

A:Cross-references: EMBL:X72816; NID:9415368; PIDN:CA51335.1; PID:94388773

C:Genetics:

A:Insertions: 19/1

C:Superfamily: Immunoglobulin V region; Immunoglobulin homology

C:Keywords: heterotetramer; Immunoglobulin

F;38-112/Domain: Immunoglobulin homology <IMM>

Query Match Best Local Similarity 100.0%; Score 26; DB 2; Length 117;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5

Db 70 LYDAK 74

## RESULT 2

H70345 conserved hypothetical protein aq\_507 - Aquifex aeolicus

C:Species: Aquifex aeolicus

C>Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 05-Nov-1999

C:Accession: H70345

R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;

Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; PMID:98196666; PMID:9537320

A:Accession: H70345

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-178 <AOF>

A:Cross-references: GB:AE000693; NID:92983148; PIDN:AA06761.1; PID:92983159; GB:AE00

C:Experimental source: strain VFS

C:Genetics:

A:Gene: aq\_507

Query Match Best Local Similarity 100.0%; Score 26; DB 2; Length 178;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5  
|||||

Db 61 LYDAK 65

### RESULT 3

B71106  
hypothetical protein PH0619 - *Pyrococcus horikoshii*

C:Species: *Pyrococcus horikoshii*  
C:Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 05-Nov-1999  
C:Accession: B71106  
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekin, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kishida, N.; Oguchi, DNA Res. 5, 55-76, 1998  
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic  
A:Reference number: A71000; MUID:98344137; PMID:9679194  
A:Accession: B71106  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-197 <KAM>  
A:Cross-references: GB:AP000003; NID:93236130; PIDN:BAA29708.1; PID:dl030651; PID:932570  
A:Experimental source: strain OT3  
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank  
C:Genetics:  
A:Gene: PH0619

Query Match 100.0%; Score 26; DB 2; Length 197;  
Best Local Similarity 100.0%; Pred. No. 62;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5  
|||||

Db 62 LYDAK 66

### RESULT 4

T24871  
hypothetical protein T1263.5 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 04-Mar-2000  
C:Accession: T24871  
R:Sim, M.  
submitted to the EMBL Data Library, January 1996  
A:Reference number: Z19945  
A:Accession: T24871  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-199 <WIL>  
A:Cross-references: EMBL:Z68752; PIDN:CMA92984.1; GSPDB:GN00022; CESP:T1263.5  
A:Experimental source: clone T1263  
C:Genetics:  
A:Gene: CESP:T1263.5  
A:Map position: 4  
A:Introns: 45/1; 166/3  
C:Superfamily: *Caenorhabditis elegans* hypothetical protein T1263.5

Query Match 100.0%; Score 26; DB 2; Length 199;  
Best Local Similarity 100.0%; Pred. No. 63;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5  
|||||

Db 134 LYDAK 138

### RESULT 5

AD1797  
Partition protein, Para homolog [imported] - *Listeria innocua* (strain Clp11262)

C:Species: *Listeria innocua*  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 17-May-2002  
C:Accession: AD1797

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurge, O.; Entian, K.D.; Fsihl, Science 294, 849-852, 2001  
A:Authors: Kref, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maltournam, A.; Ok, C.; Schueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla, A:Title: Comparative genomics of *Listeria* species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AD1797

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-253 <GL>  
A:Cross-references: GB:AU592022; PIDN:CAC98148.1; PID:916415464; GSPDB:GN00178  
A:Experimental source: strain Clp11262  
C:Genetics:  
A:Gene: para  
C:Superfamily: regulatory protein spo0

Query Match 100.0%; Score 26; DB 2; Length 253;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5  
|||||

Db 231 LYDAK 235

### RESULT 6

AF1423  
Partition protein, Para homolog para [imported] - *Listeria monocytogenes* (strain EGD-

C:Species: *Listeria monocytogenes*  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 17-May-2002  
C:Accession: AF1423  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurge, O.; Entian, K.D.; Fsihl, Science 294, 849-852, 2001  
A:Authors: Kref, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maltournam, A.; Ok, C.; Schueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla, A:Title: Comparative genomics of *Listeria* species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AF1423  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-253 <GL>  
A:Cross-references: GB:NC\_003210; PIDN:CAD01004.1; PID:916412291; GSPDB:GN00177  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: para  
C:Superfamily: regulatory protein spo0

Query Match 100.0%; Score 26; DB 2; Length 253;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5  
|||||

Db 231 LYDAK 235

### RESULT 7

G81427  
periplasmic protein Cj0111 [imported] - *Campylobacter jejuni* (strain NCCTC 11168)

C:Species: *Campylobacter jejuni*  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002  
C:Accession: G81427  
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelley, J.M.; Churcher, C.; Basham, D.; Chl, C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Bar, Nature 403, 665-668, 2000  
A:Title: The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals  
A:Reference number: AB1250; MUID:20150912; PMID:10688204  
A:Accession: G81427  
A:Status: preliminary



A:Molecule type: DNA  
 A:Residues: 1-259 <PAR>  
 A:Cross-references: GB:AL139074; GB:AL111168; MID:g6967505; PIDN:CAB72595.1; PID:g696760  
 A:Experimental source: serotype O2, strain NCTC 11168  
 C:Genetics:  
 A:Gene: Cj0111

Query Match 100.0%; Score 26; DB 2; Length 259;  
 Best Local Similarity 100.0%; Pred. No. 84;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5  
 |||||  
 DB 218 LYDAK 222

RESULT 8  
 S73522  
 Probable lipoprotein E09\_orf277 - Mycoplasma pneumoniae (strain ATCC 29342)

N:Alternate names: hypothetical protein E09\_orf277  
 C:Species: Mycoplasma pneumoniae  
 A:Variety: ATCC 29342

C:Accession: S73522  
 C:Date: 27-Feb-1997 #sequence\_revision 25-Apr-1997 #text\_change 07-Dec-1999  
 R:Hilmeirich, R.; Hilbert, H.; Plagens, H.; Parkl, E.; Li, B.C.; Herrmann, R.

Nucleic Acids Res. 24, 4420-4449, 1996  
 A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae

A:Reference number: S73522; MID:97105865; PMID:8948663  
 A:Accession: S73522  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA  
 A:Residues: 1-277 <HIM>  
 A:Cross-references: EMBL:AE000020; GB:U00089; MID:g1673852; PIDN:AB95844.1; PID:g167386

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996  
 C:Genetics:  
 A:Genetic code: SGC3  
 C:Superfamily: hypothetical protein MG440

Query Match 100.0%; Score 26; DB 2; Length 277;  
 Best Local Similarity 100.0%; Pred. No. 91;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5  
 |||||  
 DB 85 LYDAK 89

RESULT 9  
 T50351  
 Probable methyltransferase [Imported] - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe  
 C:Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 09-Jun-2000  
 C:Accession: T50351

R:Seeger, K.; Harris, D.; Lyne, M.; Rajandream, M.A.; Barrell, B.G.  
 submitted to the EMBL Data Library, February 2000  
 A:Reference number: 225064

A:Accession: T50351  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA

A:Residues: 1-281 <SEE>  
 A:Cross-references: EMBL:AL157918; PIDN:CAB76043.1; GSPDB:GND0067; SPDB:SPBC21C3.07C  
 C:Genetics:  
 A:Gene: SPDB:SPBC21C3.07C

A:Map position: 2  
 A:Introns: 44/2; 259/2

Query Match 100.0%; Score 26; DB 2; Length 281;  
 Best Local Similarity 100.0%; Pred. No. 92;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5  
 |||||

DB 164 LYDAK 168

RESULT 10  
 G97187  
 UDP-glucose 4-epimerase [Imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum  
 C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 30-Sep-2001  
 C:Accession: G97187

R:Killing, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L.  
 ; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001  
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium

A:Reference number: A96900; MID:21359325; PMID:21359325  
 A:Accession: G97187  
 A:Status: preliminary

A:Molecule type: DNA  
 A:Residues: 1-301 <KUR>  
 A:Cross-references: GB:AE001437; PIDN:AAK80290.1; PID:g15025343; GSPDB:GND0168

A:Experimental source: Clostridium acetobutylicum ATCC824  
 C:Genetics:  
 A:Gene: CA02334  
 C:Superfamily: Escherichia coli UDPglucose 4-epimerase; UDPglucose 4-epimerase homolo

Query Match 100.0%; Score 26; DB 2; Length 301;  
 Best Local Similarity 100.0%; Pred. No. 99;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5  
 |||||  
 DB 86 LYDAK 90

RESULT 11  
 C90453  
 hypothetical protein hpce-2 [Imported] - Sulfolobus solfataricus

C:Species: Sulfolobus solfataricus  
 C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 24-May-2001  
 C:Accession: C90453

R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Aways, M.J.; Ch  
 Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder

arrett, R.A.; Ragan, M.A.; Jensen, C.W.; Van der Oost, J.  
 submitted to GenBank, April 2001  
 A:Description: Sulfolobus solfataricus complete genome.

A:Reference number: A99139  
 A:Accession: C90453  
 A:Status: preliminary

A:Molecule type: DNA  
 A:Residues: 1-304 <KUR>  
 A:Cross-references: GB:AE006641; MID:g13816109; PIDN:AAK42882.1; GSPDB:GND0155

C:Genetics:  
 A:Gene: hpce-2

Query Match 100.0%; Score 26; DB 2; Length 304;  
 Best Local Similarity 100.0%; Pred. No. 102;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5  
 |||||  
 DB 35 LYDAK 39

RESULT 12  
 T15194  
 hypothetical protein F55A12.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans  
 C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 24-Nov-1999  
 C:Accession: T15194

R:Pauley, A.; Gattung, S.  
 submitted to the EMBL Data Library, May 1997  
 A:Description: The sequence of C. elegans cosmid F55A12.

A:Reference number: 218305  
 A:Accession: T15194

A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-326 <PAU>  
 A:Cross-references: EMBL:AF003130; NID:92088663; PID:92088670; PIDN:AB54128.1; GSPDB:GN  
 A:Experimental source: strain Bristol N2; clone F55A12  
 C:Genetics:  
 A:Gene: CESP:F55A12.2  
 A:Map position: 1  
 A:Introns: 39/2; 121/1; 165/3; 241/1  
 C:Superfamily: Caenorhabditis elegans hypothetical protein F55A12.2

Query Match 100.0%; Score 26; DB 2; Length 326;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5  
 |||||  
 DB 130 LYDAK 134

RESULT 13  
 B83298  
 conserved hypothetical protein PA2771 [imported] - Pseudomonas aeruginosa (strain PA01)  
 C:Species: Pseudomonas aeruginosa  
 C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C:Accession: B83298  
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lam,  
 ; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
 A:Reference number: AB2950; MUID:20437337; PMID:10984043  
 A:Accession: B83298  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-341 <STO>  
 A:Cross-references: GB:AE004705; GB:AE004091; NID:99948851; PIDN:AAG06159.1; GSPDB:GN001  
 A:Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: PA2771

Query Match 100.0%; Score 26; DB 2; Length 341;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5  
 |||||  
 DB 327 LYDAK 331

RESULT 14  
 E75266  
 GGDEF family protein - Deinococcus radiodurans (strain R1)  
 C:Species: Deinococcus radiodurans  
 C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
 C:Accession: E75266  
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
 ; M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; Mc  
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
 Science 286, 1571-1577, 1999  
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
 A:Reference number: A75250; MUID:20036896; PMID:10567266  
 A:Accession: E75266  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-356 <WHI>  
 A:Cross-references: GB:AE002079; GB:AE000513; NID:96460315; PIDN:AAF12040.1; PID:9646032  
 A:Experimental source: strain R1  
 C:Genetics:  
 A:Gene: DR2498  
 A:Map position: 1

Query Match 100.0%; Score 26; DB 2; Length 356;

Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5  
 |||||  
 DB 344 LYDAK 348

RESULT 15  
 I50679  
 CP49 protein - chicken  
 C:Species: Gallus gallus (chicken)  
 C>Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 21-Jul-2000  
 C:Accession: I50679; S58355  
 R:Sawada, K.; Agata, S.K.; Eguchi, G.; Quinlan, R.A.; Maisel, H.  
 Curr. Eye Res. 14, 545-553, 1995  
 A:Title: The predicted structure of chick lens CP49 and a variant thereof, CP49ins, th  
 A:Reference number: I50631; MUID:9606015; PMID:7587300  
 A:Accession: I50679  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-401 <SNP>  
 A:Cross-references: EMBL:X84806; NID:9953184; PIDN:CAA59271.1; PID:94377562  
 R:Agata, S.K.; Eguchi, G.; Quinlan, R.A.; Maisel, H.  
 submitted to the EMBL Data Library, February 1995  
 A:Description: The predicted structure of chick lens CP49 and a variant thereof, CP49I  
 A:Reference number: S58351  
 A:Accession: S58355  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-217, A', 219-379, 'H', 381-401 <AGA>  
 A:Cross-references: EMBL:X84806; NID:9953184; PIDN:CAA59271.1; PID:91334742  
 C:Genetics:  
 A:Gene: CP49  
 C:Superfamily: cytoskeletal keratin

Query Match 100.0%; Score 26; DB 2; Length 401;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5  
 |||||  
 DB 335 LYDAK 339

Search completed: July 11, 2003, 12:00:41  
 Job time : 3.10832 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:43:41 ; Search time 0.47817 Seconds

(without alignments)  
433.698 Million cell updates/sec

Title: US-10-091-442-7

Perfect score: 26

Sequence: 1 LYDAK 5

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SWISSPROT\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	26	100.0	277	Y44A_MYCPN	P75151 mycoplasma
2	26	100.0	380	SPB7_HUMAN	O75635 homo sapien
3	26	100.0	408	YJ95_YEAST	P47160 saccharomyc
4	26	100.0	443	MCRRB_METVA	P07956 methanococc
5	26	100.0	491	TOLC_SALEN	O54001 salmonella
6	26	100.0	543	BCHB_RHOGE	O91pas rhodocycus
7	26	100.0	571	ASNI_YEAST	P49089 saccharomyc
8	26	100.0	637	VE1_HPV35	P27220 human papil
9	26	100.0	641	NODO_RHIME	P13442 r nodg blfu
10	26	100.0	1091	NCAL_CHICK	P13590 gallus gall
11	26	100.0	1337	P152_YEAST	P35685 saccharomyc
12	26	100.0	179	YH64_SYNT3	P73042 synecocyst
13	26	100.0	257	Y418_METVA	O57861 methanococc
14	26	100.0	292	XDBB_ECOS7	O84605 escherichia
15	26	100.0	292	XDBB_ECOS7	O46800 escherichia
16	26	100.0	309	PNAD_MOUSE	O64311 mus musculu
17	26	100.0	309	PNAD_MOUSE	O28855 sus scrofa
18	26	100.0	310	DCHS_LACS3	P00862 lactobacilli
19	26	100.0	316	RLAO_PLAFB	O94660 plasmodium
20	26	100.0	358	TRMU_RICPR	O94660 plasmodium
21	26	100.0	358	TRMU_RICPR	O94660 plasmodium
22	26	100.0	370	TRMU_RICPN	O94660 plasmodium
23	26	100.0	370	TRMU_RICPN	O94660 plasmodium
24	26	100.0	377	ICED_BOVIN	O75601 bos taurus
25	26	100.0	388	P2X4_HUMAN	O99571 homo sapien
26	26	100.0	406	DHE4_PSYT1	O00128 ictalurid h
27	26	100.0	447	DHE4_PSYT1	O99571 homo sapien
28	26	100.0	471	STCH_HUMAN	P48723 aquifex pyr
29	26	100.0	501	FLAA_AOUPY	P46210 aquifex pyr
30	26	100.0	504	BCHB_ACTIRU	O94660 plasmodium
31	26	100.0	557	YOHK_BACSU	P54309 bacillus su
32	26	100.0	1084	HDA4_HUMAN	P54524 homo sapien
33	26	100.0	1178	RPOB_TREPA	O83269 treponema p
34	26	100.0	1275	COBN_PSEDE	P23929 pseudomonas

34	24	92.3	1409	1	HAP1_HAETIN	P44596 haemophilus
35	23	88.5	26	1	AMD1_CHICK	P81073 gallus gall
36	23	88.5	108	1	KVSD_MOUSE	P01636 mus musculu
37	23	88.5	185	1	ADBP_BACSU	P54570 bacillus su
38	23	88.5	191	1	Y777_METVA	O58187 methanococc
39	23	88.5	295	1	DAPF_METVA	O58189 methanococc
40	23	88.5	299	1	Y779_METVA	P75794 escherichia
41	23	88.5	308	1	PFLB_ECOS7	O00119 ictalurid h
42	23	88.5	308	1	VG53_HSV1	P04194 clostridium
43	23	88.5	319	1	DCHS_CLOPE	O57809 pyrococcus
44	23	88.5	325	1	LAID_PYROH	O67189 aquifex aeo
45	23	88.5	350	1	YB04_AOUAE	

## ALIGNMENTS

RESULT 1	Y44A_MYCPN	STANDARD:	PRT: 277 AA.
ID	Y44A_MYCPN		
AC	P75151:		
DT	01-NOV-1997 (Rel. 35, Created)		
DT	01-NOV-1997 (Rel. 35, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Hypothetical Lipoprotein MG440 homolog 1 precursor (E09_orf277).		
DN	MPN646 OR MP196.		
OS	Mycoplasma pneumoniae.		
OC	Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.		
OX	NCBI_TaxID=2104;		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=ATCC 29342 / M129;		
RX	MEDLINE=97105885; PubMed=8948633;		
RA	Himmelfreih R., Hillbert H., Plagens H., Pirkl E., Li B.-C.,		
RT	Herrmann R.;		
RT	"Complete sequence analysis of the genome of the bacterium Mycoplasma		
RL	pneumoniae.";		
CC	Nucleic Acids Res. 24:4420-4449(1996).		
CC	-1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor		
CC	(Potential).		
CC	-1- SIMILARITY: BELONGS TO THE MG439 / MG440 FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).		
CC	-----		
DR	EMBL: AE000020: AAB95844.1: -		
DR	InterPro: IPR001585; Lipoprotein_3.		
DR	Pfam: PF00938; Lipoprotein_3; 1.		
DR	ProDom: PD003276; Lipoprotein_3; 1.		
DR	PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.		
KW	Hypothetical protein; Lipoprotein; Membrane; Signal;		
KW	Complete proteome.		
FT	SIGNAL	1	25
FT	CHAIN	26	277
FT	LIPID	26	26
FT	SEQUENCE	277 AA; 31097 MW; A17DC45BED6D3065 CRC64;	
SO	SEQUENCE		
Query Match	100.0%;	Score 26;	DB 1; Length 277;
Best Local Similarity	100.0%;	Pred. No. 39;	
Matches	5;	Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	1 LYDAK 5		
Db			
	85 LYDAK 89		
RESULT 2			
SPB7_HUMAN			

```

ID  SPB7_HUMAN  STANDARD:  PRT:  380 AA.
AC  075635;
DT  15-JUN-2002 (Rel. 41, Created)
DR  15-JUN-2002 (Rel. 41, Last sequence update)
DT  15-JUN-2002 (Rel. 41, Last annotation update)
DE  Megsin (TP55) (Serpin B7).
GN  SERPINB7.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX  NCBI_TaxID=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  MEDLINE=97326116; PubMed=9182567;
RA  Tsujimoto M., Tsunooka N., Ishida N., Kurihara T., Iwasa F.,
RA  Katsuragi K., Rogi T., Kodama S., Katsuragi N., Adachi M.,
RA  Katsuragi T., Nakao M., Yamachi K., Hashino J., Haruyama M., Miura K.,
RA  Nakanishi T., Nakazato H., Teramura M., Mizoguchi H., Yamaguchi N.,
RT  "Purification, cDNA cloning, and characterization of a new serpin with
RT  megakaryocyte maturation activity."
RL  J. Biol. Chem. 272:15373-15380(1997).
RN  [2]
RP  SEQUENCE FROM N.A.
RA  TISSUE=Mesangial cells;
RA  MEDLINE=98376492; PubMed=9710452;
RA  Miyata T., Nangaku K., Suzuki D., Inagi R., Uragami K., Sakai H.,
RA  Okubo K., Kurokawa K.;
RT  "A mesangium-predominant gene, megin, is a new serpin upregulated in
RT  Iga nephropathy."
RL  J. Clin. Invest. 102:828-836(1998).
CC  -1- FUNCTION: Might function as an inhibitor of lys-specific
CC  proteases. Might influence the maturation of megakaryocytes via
CC  its action as a serpin.
CC  -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC  -1- TISSUE SPECIFICITY: Predominantly expressed in mesangial cells.
CC  -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC  or send an email to license@isb-sdb.ch).
CC  -----
DR  EMBL: D88575; BAA31232.1; -
DR  EMBL: AF027866; AAC64506.1; -
DR  HSSP: P05619; 1HE.
DR  Genew; HGNC:13902; SERPINB7.
DR  MIM: 603357; -
DR  InterPro: IPR000215; Serpin.
DR  Pfam: PF00079; serpin; 1.
DR  SMART: SM00093; SERPIN; 1.
DR  PROSITE: PS00284; SERPIN; 1.
KW  Serpin. Serine protease inhibitor.
FT  ACT_SITE 347 348 REACTIVE BOND (BY SIMILARITY).
SQ  SEQUENCE 380 AA; 42904 MW; 9A2CDB6C63CF605 CRC64;
Query Match 100.0%; Score 26; DB 1; Length 380;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 LYDAK 5
DB 121 LYDAK 125

```

```

DT  01-OCT-1996 (Rel. 34, Last annotation update)
DE  Hypothetical 45.1 kDa protein in RP55-ZMS1 intergenic region.
GN  YJR125C OR J2048.
OS  Saccharomyces cerevisiae (Baker's Yeast).
OC  Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC  Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX  NCBI_TaxID=4932;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Rose M., Koetter P., Entian K.D.;
RL  Submitted (SEP-1995) to the EMBL/Genbank/DBJ databases.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC  or send an email to license@isb-sdb.ch).
CC  -----
DR  EMBL: Z49625; CAA89656.1; -
DR  SGD: S0003886; YJR125C.
DR  InterPro: IPR001026; ENTH.
DR  Pfam: PF01417; ENTH; 1.
DR  SMART: SM00273; ENTH; 1.
KW  Hypothetical protein.
SQ  SEQUENCE 408 AA; 45091 MW; D666ECB1C0D074FE CRC64;
Query Match 100.0%; Score 26; DB 1; Length 408;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 LYDAK 5
DB 12 LYDAK 16

```

RESULT 4

```

MCRB_METVA
ID  MCRB_METVA  STANDARD:  PRT:  443 AA.
AC  P07856;
DT  01-AUG-1988 (Rel. 08, Created)
DT  01-AUG-1988 (Rel. 08, Last sequence update)
DT  15-JUN-2002 (Rel. 41, Last annotation update)
DE  Methyl-coenzyme M reductase beta subunit (EC 1.8.-.-).
GN  MCRB.
OS  Methanococcus vannielii.
OC  Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC  Methanococcaceae; Methanococcus.
OX  NCBI_TaxID=2187;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Cram D.S., Sherf B.A., Libby R.T., Mattaliano R.J., Ramachandran K.L.,
RA  Reeve J.N.;
RT  "Structure and expression of the genes, mcrBDCa, which encode the
RT  subunits of component C of methyl coenzyme M reductase in
RT  Methanococcus vannielii."
RL  Proc. Natl. Acad. Sci. U.S.A. 84:3992-3996(1987).
CC  -1- FUNCTION: Reduction of methyl-coenzyme M (2-(methylthio)
CC  ethanesulfonic acid) with 7-mercaptoheptanoylthreonine phosphate
CC  to methane and an heterodisulfide.
CC  -1- CATALYTIC ACTIVITY: CH(3)-S-CoM + H-S-HTP = CH(4) + CoM-S-S-HTP.
CC  -1- COFACTOR: THE ENZYME COMPLEX BINDS TIGHTLY (BUT NOT COVALENTLY)
CC  TO TWO MOLECULES OF COENZYME F430. F430 IS A YELLOW NICKEL
CC  PORPHINOID.
CC  -1- PATHWAY: Methanogenesis; last step.
CC  -1- SUBUNIT: HEXAMER OF TWO ALPHA, TWO BETA, AND TWO GAMMA CHAINS.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial

```

entitles requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: M16893; AAA72594.1; -

DR PIR: A27793; A27793.

DR HSSP: P11560; 1MRO.

DR InterPro: IPR003179; MCR\_beta.

DR Pfam: PF02241; MCR\_beta.1.

DR Pfam: PF02783; MCR\_beta.N.1.

KM Metanogenesis; Oxidoreductase.

SQ SEQUENCE 443 AA; 46954 MW; 7646A0D427A98517 CRC64;

Query Match 100.0%; Score 26; DB 1; Length 443;  
Best Local Similarity 100.0%; Pred. No. 64;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5  
|||||  
DB 10 LYDAK 14

RESULT 5  
TOLC\_SALEN  
ID TOLC\_SALEN STANDARD; PRT; 491 AA.  
AC 054001;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE Outer membrane protein tolC precursor.  
GN TOLC.  
OS Salmonella enteritidis.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
OX NCBI\_TaxID=592;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CD5;  
RX MEDLINE=96111490; PubMed=8801424;  
RA Stone B.J., Miller V.L.;  
RT Salmonella enteritidis has a homologue of tolC that is required for virulence in BALB/c mice."  
RL Microbiol. 17:701-712(1995).  
CC -1- FUNCTION: REQUIRED FOR PROPER EXPRESSION OF OUTER MEMBRANE PROTEIN GENES SUCH AS OMP, NMPC, PROTEIN 2, HEMOLYSIN, COLICIN V, OR COLICIN E1. MAY BE SPECIALIZED FOR SIGNAL SEQUENCE INDEPENDENT, EXTRACELLULAR SECRETION IN GRAM-NEGATIVE BACTERIA.  
CC -1- SUBCELLULAR LOCATION: Outer membrane.  
CC -1- SIMILARITY: BELONGS TO THE PRTF FAMILY OF SECRETION PROTEINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: U25178; AAC43973.1; -

DR HSSP: P02930; 1EK9.

DR InterPro: IPR003423; OEP.

DR Pfam: PF02321; OEP.2.

KM Transport; Outer membrane; Signal.

FT SIGNAL 1 24 BY SIMILARITY.  
FT CHAIN 25 491 OUTER MEMBRANE PROTEIN TOLC.  
SQ SEQUENCE 491 AA; 53725 MW; 58F0FB4C0257FA4 CRC64;

Query Match 100.0%; Score 26; DB 1; Length 491;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5  
|||||  
DB 400 LYDAK 404

RESULT 6  
BCHB\_RHOGE  
ID BCHB\_RHOGE STANDARD; PRT; 543 AA.  
AC 09UPA3;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Light-independent protochlorophyllide reductase subunit B (EC 1.18.-.-) (Li-POR subunit B).  
GN BCHB.  
OS Rhodocyclus gelatinosus (Rhodopseudomonas gelatinosa).  
OC Bacteria; Proteobacteria; beta subdivision; Comamonadaceae;  
OC Rhodovivax.  
OX NCBI\_TaxID=28068;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=IL144;  
RX MEDLINE=21240759; PubMed=11343129;  
RA Igarashi N., Harada J., Nagashima S., Matsuura K., Shinada K., Nagashima K.V.;  
RT "Horizontal transfer of the photosynthesis gene cluster and operon rearrangement in purple bacteria."  
RL J. Mol. Evol. 52:333-341(2001).  
CC -1- FUNCTION: Uses Mg-ATP and reduced ferredoxin to reduce ring D of protochlorophyllide (pchlide) to form chlorophyllide a (Chlide) (By similarity). This reaction is light-independent.  
CC -1- PATHWAY: Light-independent bacteriochlorophyll biosynthesis.  
CC -1- SUBUNIT: Protochlorophyllide reductase is thought to be composed of three subunits: bchl, bchm and bchb. Could form a heterotrimer of two bchl and two bchm subunits.  
CC -1- SIMILARITY: BELONGS TO THE CHLB / BCHB / BCHF FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AB034704; BAA94058.1; -

DR InterPro: IPR000510; Oxired\_nitrognasel.

DR Pfam: PF00148; Oxired\_nitro; 1.

KM TIGRPFAM: TIGR01278; DPOR\_BchB; 1.

DR Oxidoreductase; Photosynthesis; Bacteriochlorophyll biosynthesis.

SQ SEQUENCE 543 AA; 58697 MW; 76ADF68BA2C1D067 CRC64;

Query Match 100.0%; Score 26; DB 1; Length 543;  
Best Local Similarity 100.0%; Pred. No. 79;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5  
|||||  
DB 534 LYDAK 538

RESULT 7  
ASNI\_YEAST  
ID ASNI\_YEAST STANDARD; PRT; 571 AA.  
AC P49089;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Asparagine synthetase [glutamine-hydrolyzing] 1 (EC 6.3.5.4)  
DE (Glutamine-dependent asparagine synthetase 1).  
GN ASNI OR YPR145W OR P4659.3.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]

```

RP SEQUENCE FROM N.A.
RX MEDLINE-97109535; PubMed-8951815;
RA Dang V.D., Valens M., Bolotin-Fukuhara M., Daigman-Fornier B.;
RT "Cloning of the ASN1 and ASN2 genes encoding asparagine synthetases
RT in Saccharomyces cerevisiae: differential regulation by the
RT CCAT-box-binding factor.";
RL Mol. Microbiol. 22:681-692(1996).
RN (12)
RP SEQUENCE FROM N.A.
RC STRAIN-5288C / AB972;
RX MEDLINE-97131321; PubMed-9169875;
RA Bussey H., Storms R.K., Ahmed A., Albertmann K., Allen E., Ansoorge W.,
RA Ararajo R., Aparicio A., Barrett B.G., Badcock K., Benes V.,
RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
RA Dietrich F.S., Delius H., Dipaolo T., Dubois E., Duesterhoeft A.,
RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
RA Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.,
RA Hunnicke-Smith S., Hyman R., Johnston M., Kahan S., Kline K.,
RA Komp C., Kurd O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
RA Marathe R., Messenguy F., Mewes H.-W., Mitalpali S., Moestl D.,
RA Mueller-Auer S., Namath A., Newtlich U., Oeffner P., Pearson D.,
RA Petel F.X., Pohl T.M., Purnelle D., Schaefer M., Scharfe M.,
RA Scherens B., Schramm S., Schroeder M., Sidic A.M., Tettelin H.,
RA Urrestarazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,
RA Walsh S.V., Wambolt R., Wang Y., Wedler E., Wedler H., Winnett E.,
RA Zhong W.W., Zollner A., Vo D.H., Hani J.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";
RL Nature 387:103-105(1997).
CC -1- CANALYTIC ACTIVITY: ATP + L-aspartate + L-glutamine = AMP +
CC -1- phosphate + L-asparagine + L-glutamate.
CC -1- PATHWAY: Asparagine biosynthesis.
CC -1- SIMILARITY: CONTAINS 1 TYPE-2 GLUTAMINE AMIDOTRANSFERASE DOMAIN.
CC -1- SIMILARITY: TO OTHER ASN SYNTHETASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z48675; CAB8594.1; -
DR DR EMBL: U40829; AAB68284.1; -
DR HSP: P22106; ICT9.
DR SGD: S0006349; ASN1.
DR InterPro: IPR001962; Asn_synthase.
DR InterPro: IPR000583; GATase_2.
DR Pfam: PF00103; GATase_2; 1.
DR Pfam: PF00733; Asn_synthase; 1.
DR PROSITE: PS00443; GATASE_TYPE_II; 1.
KW Ligase: Asparagine biosynthesis; Glutamine amidotransferase;
KW Multigene family.
FT INIT_MET 0 0 BY SIMILARITY.
FT ACT_SITE 1 1 GATASE (BY SIMILARITY).
SQ SEQUENCE 571 AA; 64339 MW; 1996772387502E40 CRC64;
Query Match 100.0%; Score 26; DB 1; Length 571;
Best Local Similarity 100.0%; Pred. NO. 83;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LYDAR 5
DB 123 LYDAR 127

```

```

DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Replication protein E1.
GN E1.
OS Human papillomavirus type 35.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_Taxid=10587;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE-94265501; PubMed-8205838;
RA Dellus H., Hofmann B.;
RT "Primer-directed sequencing of human papillomavirus types.";
RL Curr. Top. Microbiol. Immunol. 186:13-31(1994).
RN (12)
RP SEQUENCE FROM N.A.
RX MEDLINE-92124753; PubMed-1310198;
RA Marich J.E., Ponteler A.V., Rice S.M., McGraw K.A., Dubensky T.W.;
RT "The phylogenetic relationship and complete nucleotide sequence of
RT human papillomavirus type 35.";
RL Virology 186:770-776(1992).
CC -1- FUNCTION: ATP-DEPENDENT DNA HELICASE REQUIRED FOR INITIATION OF
CC VIRAL DNA REPLICATION. IT FORMS A COMPLEX WITH THE VIRAL E2
CC PROTEIN. THE E1-E2 COMPLEX BINDS TO THE REPLICATION ORIGIN WHICH
CC CONTAINS BINDING SITES FOR BOTH PROTEINS.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X74477; AAA52563.1; -
DR EMBL: M74117; AAA6968.1; -
DR PIR: A40824; WIML35.
DR PIR: S36523; S36523.
DR InterPro: IPR001177; Papillom_E1.
DR Pfam: PF00519; E1; 1.
DR Pfam: PF00524; E1_N; 1.
KW Early protein; DNA replication; Helicase; ATP-binding; DNA-binding;
KW Nuclear protein.
FT NP_BIND 463 470
FT NP_BIND 31 34
FT CONFICT 142 143
FT CONFICT 235 258
FT CONFICT 269 281
FT CONFICT 417 418
FT CONFICT 515 552
FT CONFICT 587 587
FT CONFICT 613 613
FT CONFICT 72122 MW; 3C694D4451791003 CRC64;
SQ SEQUENCE 637 AA; 72122 MW; 3C694D4451791003 CRC64;
Query Match 100.0%; Score 26; DB 1; Length 637;
Best Local Similarity 100.0%; Pred. NO. 94;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LYDAR 5
DB 499 LYDAR 503

```

```

RESULT 8
ID VE1_HPV35 STANDARD: PRT: 637 AA.
AC P27220;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)

```

```

RESULT 9
ID NODO_RHIME STANDARD: PRT: 641 AA.
AC P13442;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)

```

15-JUN-2002 (Rel. 41, last annotation update)  
 DE NodQ bifunctional enzyme (Nodulation protein O) [Includes: Sulfate  
 DE adenylyltransferase subunit 1 (EC 2.7.7.4) (Sulfate adenylylate  
 DE transferase) (SAT) (ATP-sulfurylase large subunit); Adenylylsulfate  
 DE kinase (EC 2.7.1.35) (APS kinase) (ATP adenosine-5'-phosphosulfate 3'-  
 DE phosphotransferase)].  
 GN NODQ OR RA0469 OR SMA0857.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OC plasmid pSymba (megaplasmid 1).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Sinorhizobium.  
 ON NCBI\_TaxID=382;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RCR2011 / SU47;  
 RX MEDLINE=89313304; PubMed=2546009;  
 RA Cervantes E., Sharma S.B., Maillet F., Vasse J., Truchet G.,  
 RA Rosenberg C.;  
 RT "The Rhizobium meliloti host range nodQ gene encodes a protein which  
 RT shares homology with translation elongation and initiation factors.";  
 RL Mol. Microbiol. 3:745-755(1989).  
 RN [12]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92288403; PubMed=2520820;  
 RA Schmedock J., Long S.R.;  
 RT "Nucleotide sequence and protein products of two new nodulation genes  
 RT of Rhizobium meliloti, nodP and nodQ.";  
 RL Mol. Plant Microbe Interact. 2:181-194(1989).  
 RN [13]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1021;  
 RX MEDLINE=21396509; PubMed=11481432;  
 RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,  
 RA Barloy-Hubier F., Bowser L., Capela D., Galibert F., Gouzy J.,  
 RA Gurjal M., Hong A., Huizer L., Hyman R.W., Kahn D., Kahn M.L.,  
 RA Kaiman S., Keating D.H., Palm C., Beck M.C., Surzycki R., Wells D.H.,  
 RA Kahn K.-C., Davis R.W., Federspiel N.A., Long S.R.;  
 RT "Nucleotide sequence and predicted functions of the entire  
 RT Sinorhizobium meliloti pSymba megaplasmid.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).  
 CC -1- FUNCTION: PROPOSED TO PROVIDE ACTIVATED SULFATE FOR TRANSFER TO  
 CC NOD FACTOR. ATP SULFURYLASE MAY BE THE GTPASE, REGULATING ATP  
 CC SULFURYLASE ACTIVITY (BY SIMILARITY).  
 CC -1- FUNCTION: APS KINASE CATALYZES THE SYNTHESIS OF ACTIVATED SULFATE  
 CC (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: ATP + sulfate -> diphosphate + adenylylsulfate.  
 CC -1- CATALYTIC ACTIVITY: ATP + adenylylsulfate -> ADP + 3'-  
 CC phosphoadenylylsulfate.  
 CC -1- SUBUNIT: SULFATE-ACTIVATING ENZYMES, NODP AND NODO, MAY BE  
 CC PHYSICALLY ASSOCIATED (POTENTIAL).  
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO THE GTP-BINDING  
 CC ELONGATION FACTOR FAMILY. CTSN/NODO SUBFAMILY.  
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION, BELONGS TO THE APS KINASE  
 CC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: X14809; CAA32914.1; -;  
 DR EMBL: M68858; AAA26343.1; -;  
 DR EMBL: AE007237; AAK65127.1; -;  
 DR PIR: S14899; 222RNO.  
 DR InterPro: IPR002891; APS\_kinase.  
 DR InterPro: IPR004161; EFTU\_D2.  
 DR InterPro: IPR000795; EF\_GTPbind.  
 DR InterPro: IPR005225; Small\_GTP.  
 DR Pfam: PF00009; GTP\_EFTU; 1.  
 DR Pfam: PF01583; APS\_kinase; 1.

Pfam: PF03144; GTP\_EFTU\_D2; 1.  
 DR ProDom: PD002350; APS\_kinase; 1.  
 DR TIGRFAMs: TIGR00231; small\_gtp; 1.  
 DR TIGRFAMs: TIGR00455; apsk; 1.  
 DR PROSITE: PS00301; EFATOR\_GTP; 1.  
 KW Plasmid; Nodulation; Transferase; Nucleotidyltransferase; GTP-binding;  
 KM Kinase; ATP-binding; Multifunctional enzyme; Complete proteome.  
 FT DOMAIN 1 458  
 FT DOMAIN 459 641  
 FT NE\_BIND 31 38  
 FT NE\_BIND 110 114  
 FT NE\_BIND 165 168  
 FT NE\_BIND 467 474  
 FT ACT\_SITE 524 524  
 FT ACT\_SITE 524 524  
 FT SEQUENCE 641 AA; 70614 MW; 1E1261F04ED33A93 CRC64;  
 SO  
 Query Match 100.0%; Score 26; DB 1; Length 641;  
 Best Local Similarity 100.0%; Pred. No. 94;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LYDAK 5  
 Db 45 LYDAK 49  
 RESULT 10  
 NCAL\_CHICK STANDARD; PRT; 1091 AA.  
 ID NCAL\_CHICK  
 AC P13590; Q90919; Q90918;  
 DT 01-JUN-1998 (Rel. 13, Created)  
 DT 15-JUN-1998 (Rel. 36, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Neural cell adhesion molecule 1, 180 kDa isoform precursor (N-CAM  
 DE 180).  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE OF 1-175 FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE=87206190; PubMed=3576199;  
 RA Cunningham B.A., Hemperly J.J., Murray B.A., Prediger E.A.,  
 RA Brackenbury R., Edelman G.M.;  
 RT "Neural cell adhesion molecule: structure, immunoglobulin-like  
 RT domains, cell surface modulation, and alternative RNA splicing.";  
 RL Science 236:799-806(1987).  
 RN [2]  
 RP SEQUENCE OF 128-1091 FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE=86206089; PubMed=3458261;  
 RA Hemperly J.J., Murray B.A., Edelman G.M., Cunningham B.A.;  
 RT "Sequence of a cDNA clone encoding the polysialic acid-rich and  
 RT cytoplasmic domains of the neural cell adhesion molecule N-CAM.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 83:3037-3041(1986).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM D).  
 RX MEDLINE=87092340; PubMed=3467341;  
 RA Hemperly J.J., Edelman G.M., Cunningham B.A.;  
 RT "cDNA clones of the neural cell adhesion molecule (N-CAM) lacking a  
 RT membrane-spanning region consistent with evidence for membrane  
 RT attachment via a phosphatidylinositol intermediate.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 83:9822-9826(1986).  
 RN [4]  
 RP SEQUENCE OF 810-1069 FROM N.A.  
 RX MEDLINE=87033934; PubMed=3771645;  
 RA Murray B.A., Owens G.C., Prediger E.A., Crossin K.L.,  
 RA Cunningham B.A., Edelman G.M.;  
 RT "Cell surface modulation of the neural cell adhesion molecule  
 RT resulting from alternative mRNA splicing in a tissue-specific  
 RT developmental sequence.";  
 RL J. Cell Biol. 103:1431-1439(1986).  
 RN [5]

RP SEQUENCE OF 1-17 FROM N.A.  
 RX MEDLINE-9312797; PubMed-1478668;  
 RA Colwell G., Li B., Forrest D., Brackenbury R.;  
 RT "Conserved regulatory elements in the promoter region of the N-CAM  
 gene";  
 RL Genomics 14:875-882(1992).  
 RP [6]  
 RP SEQUENCE OF 1-17 FROM N.A.  
 RC STRAIN-White leghorn; TISSUE-Erythrocyte;  
 RA Sasser M., Covault J.;  
 RL Submitted (Aug-1993) to the EMBL/GenBank/DBJ databases.  
 CC -I- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN  
 CC NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF  
 CC NEURITES, ETC.  
 CC -I- ALTERNATIVE PRODUCTS: 4 isoforms; A/N-CAM 180 (shown here), B/N-  
 CC CAM 140, C and D; are produced by alternative splicing.  
 CC -I- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.  
 CC -I- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.  
 CC -I- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: M15861; AAB59958.1; -  
 DR EMBL: M15860; AAB59958.1; JOINED.  
 DR EMBL: M15922; AAB59958.1; JOINED.  
 DR EMBL: M15923; AAB59958.1; JOINED.  
 DR EMBL: M15924; AAB59958.1; JOINED.  
 DR EMBL: M21178; AAB59958.1; JOINED.  
 DR EMBL: M21179; AAB59958.1; JOINED.  
 DR EMBL: M21180; AAB59958.1; JOINED.  
 DR EMBL: M15929; AAB59958.1; JOINED.  
 DR EMBL: M15929; AAB59958.1; JOINED.  
 DR EMBL: M15930; AAB59958.1; JOINED.  
 DR EMBL: M15931; AAB59958.1; JOINED.  
 DR EMBL: M15932; AAB59958.1; JOINED.  
 DR EMBL: M15933; AAB59958.1; JOINED.  
 DR EMBL: M15934; AAB59958.1; JOINED.  
 DR EMBL: L29437; AAB59958.1; JOINED.  
 DR EMBL: M15935; AAB59958.1; JOINED.  
 DR EMBL: M15937; AAB59958.1; JOINED.  
 DR EMBL: M15938; AAB59958.1; JOINED.  
 DR EMBL: M15939; AAB59958.1; JOINED.  
 DR EMBL: M15861; AAB59959.1; -  
 DR EMBL: M15860; AAB59959.1; JOINED.  
 DR EMBL: M15922; AAB59959.1; JOINED.  
 DR EMBL: M15923; AAB59959.1; JOINED.  
 DR EMBL: M15924; AAB59959.1; JOINED.  
 DR EMBL: M21178; AAB59959.1; JOINED.  
 DR EMBL: M21179; AAB59959.1; JOINED.  
 DR EMBL: M21180; AAB59959.1; JOINED.  
 DR EMBL: M15929; AAB59959.1; JOINED.  
 DR EMBL: M15930; AAB59959.1; JOINED.  
 DR EMBL: M15931; AAB59959.1; JOINED.  
 DR EMBL: M15932; AAB59959.1; JOINED.  
 DR EMBL: M15933; AAB59959.1; JOINED.  
 DR EMBL: L29437; AAB59959.1; JOINED.  
 DR EMBL: M15935; AAB59959.1; JOINED.  
 DR EMBL: M15937; AAB59959.1; JOINED.  
 DR EMBL: M15939; AAB59959.1; JOINED.  
 DR EMBL: M15936; AAB59957.1; -  
 DR EMBL: M15860; AAB59957.1; JOINED.  
 DR EMBL: M15922; AAB59957.1; JOINED.  
 DR EMBL: M15923; AAB59957.1; JOINED.  
 DR EMBL: M15924; AAB59957.1; JOINED.  
 DR EMBL: M21178; AAB59957.1; JOINED.  
 DR EMBL: M21179; AAB59957.1; JOINED.  
 DR EMBL: M21180; AAB59957.1; JOINED.  
 DR EMBL: M15929; AAB59957.1; JOINED.

DR EMBL: M15930; AAB59957.1; JOINED.  
 DR EMBL: M15931; AAB59957.1; JOINED.  
 DR EMBL: M15932; AAB59957.1; JOINED.  
 DR EMBL: M15933; AAB59957.1; JOINED.  
 DR EMBL: M15934; AAB59957.1; JOINED.  
 DR EMBL: L29437; AAB59957.1; JOINED.  
 DR EMBL: M15935; AAB59957.1; JOINED.  
 DR EMBL: M15936; AAB59957.1; JOINED.  
 DR EMBL: X04479; CAB51638.1; -  
 DR EMBL: X70342; CAA49807.1; -  
 DR EMBL: Z12128; CAA78113.1; -  
 DR PIR: A43613; ICHNL.  
 DR HSSP: P40189; 1BDH.  
 DR InterPro: IPR003961; FN\_III.  
 DR InterPro: IPR003006; Ig\_MHC.  
 DR InterPro: IPR003598; Ig\_C2.  
 DR InterPro: IPR003600; Ig\_1like.  
 DR Pfam: PF00041; f03; 2.  
 DR Pfam: PF00047; f9; 5.  
 DR SMART: SM00060; FN3; 2.  
 DR SMART: SM00410; Ig\_1like; 1.  
 DR SMART: SM00408; ICG2; 4.  
 DR Cell adhesion: Glycoprotein: Transmembrane; Repeat;  
 KW Immunoglobulin domain: Alternative splicing; Signal.  
 FT SIGNAL 1 19  
 FT CHAIN 20 1091  
 FT FT  
 FT DOMAIN 20 711  
 FT TRANSMEM 712 729  
 FT DOMAIN 730 1091  
 FT FT  
 FT DOMAIN 34 103  
 FT DOMAIN 132 103  
 FT DOMAIN 132 103  
 FT DOMAIN 228 294  
 FT DOMAIN 322 392  
 FT DOMAIN 419 486  
 FT DOMAIN 518 595  
 FT DOMAIN 624 682  
 FT DOMAIN 152 156  
 FT DOMAIN 161 165  
 FT DISULFID 41 96  
 FT DISULFID 139 189  
 FT DISULFID 235 287  
 FT DISULFID 329 385  
 FT DISULFID 426 479  
 FT CARBOHYD 222 222  
 FT CARBOHYD 315 315  
 FT CARBOHYD 347 347  
 FT CARBOHYD 423 423  
 FT CARBOHYD 449 449  
 FT CARBOHYD 478 478  
 FT CARSPLIC 702 726  
 FT FT  
 FT VARSPLIC 727 1091  
 FT VARSPLIC 771 809  
 FT VARSPLIC 810 1070  
 FT SEQUENCE 1091 AA; 117415 MW; B151367002DF88BD CRC64;  
 SQ  
 Query Match 100.0%; Score 26; DB 1; Length 1091;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LYDAK 5  
 DB 549 LYDAK 553  
 RESULT 11  
 ID P152.YEAST  
 AC P39685;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Nuclear envelope pore membrane protein POM152 (P150).  
 GN POM152 OR YMR129W OR YW5553.05.



```

OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-W303;
RX MEDLINE=94186543; PubMed=8138573;
RA WOZNIAK R.W., BLODEL G., ROUT M.P.;
RT POM152 is an integral protein of the pore membrane domain of the
RT yeast nuclear envelope.
RL Cell Biol. 125:31-42(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA BADCOCK K., CHURCHER C., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: COMPONENT OF THE NUCLEAR PORE COMPLEX WHICH MAY PROVIDE
CC THE NECESSARY ASYMMETRY REQUIRED FOR ANCHORING STRUCTURES SUCH AS
CC CYTOPLASMIC FILAMENTS AND THE NUCLEOPLASMIC CAGE.
CC -1- SUBUNIT: INTERACTS WITH NUP170 AND/OR NUP188.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. NUCLEAR PORE
CC MEMBRANE.
CC -1- PTM: THE N-TERMINUS IS BLOCKED.
CC -1- SIMILARITY: TO S.POMBE SPEC29A10.07.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z31592; CAA83469.1; -
DR EMBL; Z48622; CAA88554.1; -
DR PIR; A53824; A53824.
DR SGD; S0004736; POM152.
KM Nuclear protein; Transmembrane; Repeat; Glycoprotein.
FT DOMAIN 1 175 PORE SIDE (POTENTIAL).
FT TRANSMEM 176 195 POTENTIAL.
FT DOMAIN 196 1337 CISTERNA SIDE (POTENTIAL).
FT CARBOHYD 280 280 N-LINKED (GLCNAC...);
FT DOMAIN 390 1276 8 X 24 AA APPROXIMATE REPEATS.
FT REPEAT 390 413 1.
FT REPEAT 626 650 2.
FT REPEAT 732 755 3.
FT REPEAT 836 859 4.
FT REPEAT 943 966 5.
FT REPEAT 1058 1077 6.
FT REPEAT 1157 1178 7.
FT REPEAT 1253 1276 8.
SQ SEQUENCE 1337 AA; 151651 MW; A024A42069193898 CRC64;

Query Match 100.0%; Score 26; DB 1; Length 1337;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA MIYAJIMA N., MIYOSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA HOSONUCHI T., MATSUO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,
RA SHIMPO S., TAKEUCHI C., WADA T., MATANABE A., YAMADA M., YASUDA M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -1- SIMILARITY: BELONGS TO THE CAPAB / TERDEZ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D90903; BAA17063.1; -
DR InterPro; IPR003325; TerD.
DR Pfam; PF02342; TerD; 1.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 179 AA; 19422 MW; 19BC13A5C710FEA7 CRC64;

Query Match 92.3%; Score 24; DB 1; Length 179;
Best Local Similarity 80.0%; Pred. No. 77;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 LYDAK 5
Db 662 LYDAK 666

```

```

RESULT 13
ID Y418_METJA STANDARD; PRT; 257 AA.
AC 057861;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ0418.
GN MJ0418.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA BILT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,
RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,
RA KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.T.,
RA OVERBECK R., KIRKNESS E.F., WEINSTOCK K.G., MERICK J.M., GLODOK A.,
RA SCOTT J.L., GEORGHAGEN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,
RA UTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,
RA COTTON M.D., ROBERTS K.M., HURST M.A., KALNE B.P., BORODOVSKY M.,
RA KIENK H.-P., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entties requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U67494; AAB98416.1; -  
 KM Hypothetical protein; Transmembrane; Complete proteome.  
 FT TRANSMEM 7 POTENTIAL.  
 SQ SEQUENCE 257 AA; 28765 MW; CFEAIAA6AB38000 CRC64;

Query Match  
 Best Local Similarity 92.3%; Score 24; DB 1; Length 257;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 LYDAK 5  
 Db 82 IYDAK 86

RESULT 14  
 XDBH\_ECO57 STANDARD; PRT; 292 AA.

AC 08X65;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DE Xanthine dehydrogenase, FAD binding subunit (EC 1.1.1.204).  
 GN XDBH OR 24206 OR ECS3740.  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia  
 OX NCBI\_TaxID=83334;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G., Burland V., Mau B., Glasner J.D.,  
 Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,  
 Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 Welch R.A., Blattner F.R.;  
 RA "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";  
 RT Nature 409:529-533(2001).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=O157:H7 / RIMD 0509952;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tode T.,  
 Ikeda T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,  
 Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
 RA "Complete genome sequence of enterohaemorrhagic Escherichia coli  
 O157:H7 and genomic comparison with a laboratory strain K-12.";  
 RT DNA Res. 8:11-22(2001).  
 RL -1- FUNCTION: Presumed to be a dehydrogenase, but possibly an oxidase.  
 Participates in limited purine salvage (requires aspartate) but  
 does not support aerobic growth on purines as the sole carbon  
 source (purine catabolism) (By similarity).  
 CC -1- CATALYTIC ACTIVITY: Xanthine + NAD(+) + H(2)O = urate + NADH.  
 CC -1- COFACTOR: FAD (By similarity).  
 CC -1- PATHWAY: Purine catabolism; first committed step.  
 CC -1- SUBUNIT: Heterotrimer of xdbh, xdbb and xdcn (Probable).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AE005516; AAG57996.1; -  
 DR EMBL: AP002563; BAB37163.1; -;

KW Oxidoreductase; NAD; Flavoprotein; FAD; Purine metabolism;  
 KM Purine salvage; Complete proteome.  
 SQ SEQUENCE 292 AA; 31561 MW; FEC44F990BF9BC1 CRC64;

Query Match  
 Best Local Similarity 92.3%; Score 24; DB 1; Length 292;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 LYDAK 5  
 Db 130 IYDAK 134

RESULT 15  
 XDBH\_ECOLI STANDARD; PRT; 292 AA.

AC 046800;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DE 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Xanthine dehydrogenase, FAD binding subunit (EC 1.1.1.204).  
 GN XDBH OR B2867.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G., III, Bloch C.A., Perna N.T., Burland V.,  
 Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 Mau B., Shao Y.;  
 RA "The complete genome sequence of Escherichia coli K-12.";  
 RT Science 277:1453-1474(1997).  
 RL [2]  
 RN DISCUSSION OF FUNCTION.  
 RC STRAIN=K12 / W3110;  
 RX MEDLINE=20444178; PubMed=10986234;  
 RA Xi H., Schneider B.L., Reitzer L.;  
 RA "Purine catabolism in Escherichia coli and function of xanthine  
 dehydrogenase in purine salvage.";  
 RT J. Bacteriol. 182:5332-5341(2000).  
 RL -1- FUNCTION: Presumed to be a dehydrogenase, but possibly an oxidase.  
 Participates in limited purine salvage (requires aspartate) but  
 does not support aerobic growth on purines as the sole carbon  
 source (purine catabolism).  
 CC -1- CATALYTIC ACTIVITY: Xanthine + NAD(+) + H(2)O = urate + NADH.  
 CC -1- COFACTOR: FAD (By similarity).  
 CC -1- PATHWAY: Purine catabolism; first committed step.  
 CC -1- SUBUNIT: Heterotrimer of xdbh, xdbb and xdcn (Probable).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U28375; AAB83048.1; -  
 DR EMBL: AE000370; AAC75905.1; -  
 DR Ecogene; EGI3050; xdbb.  
 DR InterPro: IPR005107; CO\_deh\_flav\_C.  
 DR InterPro: IPR002346; dehydrotrog\_molyb.  
 DR Pfam: PF00941; FAD\_binding\_5; 1.  
 DR Pfam: PF03450; CO\_deh\_flav\_C; 1.  
 KW Oxidoreductase; NAD; Flavoprotein; FAD; Purine metabolism;  
 KM Purine salvage; Complete proteome.  
 SQ SEQUENCE 292 AA; 31556 MW; 74A07D137DA857E8 CRC64;

Query Match 92.3%; Score 24; DB 1; Length 292;

Best Local Similarity 80.0%; Pred. No. 1.3e+02;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5  
:||||  
Db 130 TYDAK 134

Search completed: July 11, 2003, 11:55:18  
Job time : 1.47817 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:44:31 ; Search time 2.18295 Seconds  
(without alignments)  
471.946 Million cell updates/sec

Title: US-10-091-442-7  
Perfect score: 26  
Sequence: 1 LYDAK 5

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues  
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPREMBL\_21: \*  
1: sp.archaea: \*  
2: sp.bacteria: \*  
3: sp.fungi: \*  
4: sp.human: \*  
5: sp.invertebrate: \*  
6: sp.mammal: \*  
7: sp.mhc: \*  
8: sp.organelle: \*  
9: sp.phage: \*  
10: sp.plant: \*  
11: sp rodent: \*  
12: sp.virus: \*  
13: sp.vertebrate: \*  
14: sp.unclassified: \*  
15: sp.virus: \*  
16: sp.bacteriophage: \*  
17: sp.archaeo: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	26	100.0	107	17	0973B2 sulfolobus
2	26	100.0	108	9	0971J3 lactobacilli
3	26	100.0	135	5	08SUH3 encephalito
4	26	100.0	175	5	001760 caenorhabditis
5	26	100.0	178	16	066796 aquifex
6	26	100.0	195	17	08U3H5
7	26	100.0	197	17	08U3H5
8	26	100.0	199	5	022438
9	26	100.0	213	10	09ASL1
10	26	100.0	230	12	088960
11	26	100.0	250	2	09PCV9
12	26	100.0	250	5	08T3H1
13	26	100.0	253	16	0926W7
14	26	100.0	259	16	09P3J5
15	26	100.0	281	3	09P7L6
16	26	100.0	288	10	09CA01

17	26	100.0	299	17	08TKS0 methanococcus
18	26	100.0	301	16	097GN2 clostridium
19	26	100.0	304	17	097V63 sulfolobus
20	26	100.0	306	2	0936E9 staphylococcus
21	26	100.0	329	2	0936E9 staphylococcus
22	26	100.0	331	9	094M05 bacteriophage
23	26	100.0	334	2	08RPD7 legionella
24	26	100.0	341	16	091072 pseudomonas
25	26	100.0	354	12	0966M3 turkey
26	26	100.0	354	12	067629 marek
27	26	100.0	356	16	09RRJ2 delinococcus
28	26	100.0	357	10	09FJ41 arabis
29	26	100.0	370	16	08XK23 clostridium
30	26	100.0	377	16	0985U9 rhizobium
31	26	100.0	388	2	09KX7 staphylococcus
32	26	100.0	401	13	090976 gallus
33	26	100.0	409	16	098IS9 staphylococcus
34	26	100.0	421	16	09EV35 staphylococcus
35	26	100.0	428	16	098M33 rhizobium
36	26	100.0	430	16	09R1V4 streptomyces
37	26	100.0	439	16	08U6J7 agrobacterium
38	26	100.0	448	17	0974Y6 sulfolobus
39	26	100.0	450	13	090809 gallus
40	26	100.0	460	5	08WS95 glosina
41	26	100.0	464	10	094ET0 hevea
42	26	100.0	464	10	0944F8 hevea
43	26	100.0	474	10	093773 pinus
44	26	100.0	475	5	094404 caenorhabditis
45	26	100.0	491	16	082L24 salmone

#### ALIGNMENTS

RESULT 1  
ID: 0973B2 PRELIMINARY; PRT: 107 AA.  
AC: 0973B2;  
DT: 01-DEC-2001 (TREMUREL. 19, Last sequence update)  
DE: 01-JUN-2002 (TREMUREL. 21, Last annotation update)  
DD: Putative transcriptional regulator.  
GN: SP0980.  
OS: Sulfolobus tokodaii.  
OC: Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
OC: Sulfolobus.  
OX: NCBI\_TaxID=111955;  
RN: [1]  
RP: SEQUENCE FROM N.A.  
RX: PubMed=11572475;  
RA: Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S., Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya N., Kato Y., Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Masuda S., Yanagisaki M., Nishimura M., Yamagishi A., Oshima T., Kikuchi H.;  
RT: Complete genome sequence of an aerobic thermophilic  
RT: Crenarchaeon, Sulfolobus tokodaii strain7.  
RL: DNA Res. 8:123-140(2001).  
DR: EMBL: AP000984; BAB6001.1;  
KW: Hypothetical protein; Complete proteome.  
SO: SEQUENCE 107 AA; 12386 MW; 2CF24E31B02D3C68 CRC64;

Query Match 100.0%; Score 26; DB 17; Length 107;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5  
DB 5 LYDAK 9

```

RESULT 2
Q9T1J3      PRELIMINARY;      PRT;      108 AA.
ID 09T1J3
AC 09T1J3;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Rad protein.
GN RAD.
OS Lactobacillus bacteriophage phi adh.
OC Viruses: dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX NCBI_TaxID=12417;
RN [1]
RP SEQUENCE FROM N.A.
RA Altermann E.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases..
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=9384014; PubMed=10452953;
RA Altermann E., Klein J., Henrich B.;
RT "Primary structure and features of the genome of the Lactobacillus gasseri temperate bacteriophage phi-adh.";
RL Gene 236:333-346(1999).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=95138034; PubMed=7836307;
RA Henrich B., Bishofer B., Bialesi U.;
RT "Primary structure and functional analysis of the lysis genes of Lactobacillus gasseri bacteriophage phi-adh.";
RL J. Bacteriol. 177:723-732(1995).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=93231538; PubMed=8472961;
RA Fremux C., De Antoni G., Raya R., Kleenhammer T.;
RT "Genetic organization and sequence of the region encoding integrative functions from Lactobacillus gasseri temperate bacteriophage phi-adh.";
RL Gene 126:61-66(1993).
RN [5]
RP SEQUENCE FROM N.A.
RA Engel G., Altermann E., Klein J., Henrich B.;
RT "Structure of a genome region of the Lactobacillus gasseri temperate phage phi adh covering a repressor gene and cognate promoters.";
RL Gene 210:67-70(1998).
DR EMBL: AJ131519; CAB52483.1;
DR InterPro: IPR001387; HTH_3.
DR Pfam: PF01381; HTH_3.1.
DR SMART: SM00530; HTH_XRE; 1.
SQ SEQUENCE 108 AA; 12086 MW; EC9109B3318BC5E2 CRC64;

Query Match      100.0%; Score 26; DB 9; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 LYDAK 5
Db      104 LYDAK 108

```

```

RC STRAIN=GB-M1;
RA Genoscope;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21576510; PubMed=11719806;
RA Katinka M.D., Duprat S., Cornillot E., Metenier G., Thomarat F.,
RA Prensier G., Barbe V., Peyreallade E., Brottier P., Minkler P.,
RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,
RA Weissenbach J., Vivas C.P.;
RT "Genome sequence and gene compaction of the eukaryote parasite Encephalitozoon cuniculi.";
RL Nature 414:450-453(2001).
DR EMBL: AL590449; CAD25742.1;
KM Hypothetical protein.
SQ SEQUENCE 135 AA; 16153 MW; A202A9AD801B1B5F CRC64;

Query Match      100.0%; Score 26; DB 5; Length 135;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 LYDAK 5
Db      87 LYDAK 91

RESULT 4
ID 001760      PRELIMINARY;      PRT;      175 AA.
AC 001760;
DT 01-JUL-1997 (TReMBLrel. 04, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Hypothetical 19.3 kDa protein.
GN F55A12.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA "The sequence of C. elegans cosmid F55A12.";
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF003130; AAB54128.2;
KM Hypothetical protein.
SQ SEQUENCE 175 AA; 19251 MW; 443B94949EEF2897B CRC64;

Query Match      100.0%; Score 26; DB 5; Length 175;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 LYDAK 5
Db      54 LYDAK 58

```

RESULT 5  
 066796 PRELIMINARY; PRT; 178 AA.  
 ID 066796  
 AC 066796;  
 DT 01-AUG-1998 (TReMBLrel. 07, Created)  
 DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)  
 DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)  
 DE Hypothetical protein AQ\_507.  
 GN AQ\_507.  
 OS Aquifex aeolicus.  
 OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.  
 NCBI\_TaxID=63363;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VF5;  
 RX MEDLINE=98196666; PubMed=9537320;  
 RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
 RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aujay M., Huber R.,  
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;  
 RT "The complete genome of the hyperthermophilic bacterium Aquifex  
 RT aeolicus.";  
 RL Nature 392:353-358(1998).  
 DR EMBL: AE000693; AAC06761.1; -;  
 DR InterPro: IPR002934; NTP\_transf.  
 DR InterPro: IPR001201; PAP\_25A\_core.  
 DR Pfam: PF01909; NTP\_transf.2; 1.  
 KW Hypothetical protein: Complete proteome.  
 SQ SEQUENCE 178 AA; 20421 MW; C3EF9688B9A78F42 CRC64;

Query Match 100.0%; Score 26; DB 16; Length 178;  
 Best Local Similarity 100.0%; Pred. NO. 2.4e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5  
 |||||  
 Db 61 LYDAK 65

RESULT 6  
 0803H5 PRELIMINARY; PRT; 195 AA.  
 ID 0803H5  
 AC 0803H5;  
 DT 01-JUN-2002 (TReMBLrel. 21, Created)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE Transcription initiation factor TFIIIE.  
 GN PF0491.  
 OS Pyrococcus furiosus.  
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
 OC Pyrococcus.  
 NCBI\_TaxID=2261;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VC1 / DSM 3638 / ATCC 43587 / JCM 8422;  
 RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;  
 RT "The complete sequence of the Pyrococcus furiosus genome.";  
 RL EMBL: AE010171; AAL80615.1; -;  
 DR EMBL: AE010171; AAL80615.1; -;  
 KW Initiation factor: Complete proteome.  
 SQ SEQUENCE 195 AA; 22885 MW; F7507AF9F3353763 CRC64;

Query Match 100.0%; Score 26; DB 17; Length 195;  
 Best Local Similarity 100.0%; Pred. NO. 2.4e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5  
 |||||  
 Db 58 LYDAK 62

RESULT 7  
 058353 PRELIMINARY; PRT; 197 AA.

AC 058353;  
 DT 01-AUG-1998 (TReMBLrel. 07, Created)  
 DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE Hypothetical protein PH0619.  
 GN PH0619.  
 OS Pyrococcus horikoshii.  
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
 OC Pyrococcus.  
 NCBI\_TaxID=53953;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OT3;  
 RX MEDLINE=98344137; PubMed=9679194;  
 RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,  
 RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,  
 RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,  
 RA Funahashi T., Tanaka T., Kudo Y., Yamazaki J., Kushiida N., Ouchi A.,  
 RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,  
 RA Masuchi Y., Shizuwa H., Kikuchi H.;  
 RT "Complete sequence and gene organization of the genome of a hyper-  
 RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";  
 RL DNA Res. 5:55-76(1998).  
 DR EMBL: AP000003; BAA29708.1; -;  
 DR InterPro: IPR002853; TFIIIE\_alpha.  
 DR InterPro: IPR005241; TFIIIE\_alpha-rel.  
 DR Pfam: PF02002; TFIIIE\_alpha; 1.  
 DR SMART: SM00531; TFIIIE; 1.  
 DR TIGRfams: TIGR00373; TIGR00373; 1.  
 KW Hypothetical protein: Complete proteome.  
 SQ SEQUENCE 197 AA; 23127 MW; 2A0AA7110389AF61 CRC64;

Query Match 100.0%; Score 26; DB 17; Length 197;  
 Best Local Similarity 100.0%; Pred. NO. 2.4e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5  
 |||||  
 Db 62 LYDAK 66

RESULT 8  
 022438 PRELIMINARY; PRT; 199 AA.  
 ID 022438  
 AC 022438;  
 DT 01-NOV-1996 (TReMBLrel. 01, Created)  
 DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
 DE T12G3.5 protein.  
 GN T12G3.5.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Pelodermidae; Caenorhabditis.  
 NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Sims M.A.;  
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA none;  
 RT "Genome sequence of the nematode C. elegans: A platform for  
 RT investigating biology.";  
 RL Science 282:2012-2018(1998).  
 DR EMBL: Z68752; CAA92984.1; -;  
 SQ SEQUENCE 199 AA; 23636 MW; A4FF219673154006 CRC64;

Query Match 100.0%; Score 26; DB 5; Length 199;  
 Best Local Similarity 100.0%; Pred. NO. 2.4e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5

DB 134 LYDAK 138

## RESULT 9

Q9ASL1 PRELIMINARY; PRT; 213 AA.  
ID Q9ASL1;  
AC Q9ASL1;  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)  
DT 01-MAR-2002 (TREMBLrel. 20, last annotation update)  
DE P0439B06.1 protein (P0482C06.29 protein).  
GN P0439B06.1 OR P0482C06.29.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzaceae; Oryza.  
OX NCBI\_TaxID=4530;

RM (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN-CV, NIPPONBARE;  
RA Sasaki T., Matsumoto T., Yamamoto K.;  
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC clone: P0439B06.";  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
RN (12)  
RP SEQUENCE FROM N.A.  
RC STRAIN-CV, NIPPONBARE;  
RA Sasaki T., Matsumoto T., Yamamoto K.;  
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC clone: P0482C06.";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AP002882; BAB39866.1;  
DR EMBL; AP002845; BAB78632.1;  
DR InterPro: IPR000169; SHPOC.acsite.  
DR PROSITE: PS00639; THIOX.PROTEASE\_HIS; UNKNOWN\_1.  
SQ SEQUENCE 213 AA; 22915 MW; 9EAA4F8D9184AA3 CRC64;

Query Match 100.0%; Score 26; DB 10; Length 213;  
Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5  
DB 67 LYDAK 71

## RESULT 10

Q88960 PRELIMINARY; PRT; 230 AA.

ID Q88960;  
AC Q88960;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)  
DT 01-MAR-2002 (TREMBLrel. 20, last annotation update)  
DE DNA polymerase (Fragment).  
OS Unidentified phycodnavirus clone OT05.  
OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae;  
OC Environmental samples.  
OX NCBI\_TaxID=43245;

RP SEQUENCE FROM N.A.  
RC MEDLINE=96316404; Pubmed=8702280;  
RA Chen F., Suttle C.A., Short S.M.;  
RT "Genetic diversity in marine algal virus communities as revealed by sequence analysis of DNA polymerase genes";  
RL Appl. Environ. Microbiol. 62:2869-2874(1996).  
-I- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE - N DIPHOSPHATE + (DNA)(N).

CC -I- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.  
CC EMBL; U36935; AAB18299.1;  
DR InterPro: IPR002064; DNA\_pol\_B.  
DR Pfam: PF00136; DNA\_pol\_B; 2.  
DR PRINTS; PR00106; DNAPOB.

DR PROSITE: PS00116; DNA\_POLYMERASE\_B; UNKNOWN\_1.  
KW DNA replication; DNA-binding; DNA-directed DNA polymerase.  
FT NON\_TER 1 1  
FT NON\_TER 230 230  
SQ SEQUENCE 230 AA; 25543 MW; 983319E04EC18388 CRC64;

Query Match 100.0%; Score 26; DB 12; Length 230;  
Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5  
DB 104 LYDAK 108

## RESULT 11

Q9FCV9 PRELIMINARY; PRT; 250 AA.

ID Q9FCV9;  
AC Q9FCV9;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)  
DE Para protein.  
GN PARA.  
OS Listeria monocytogenes.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
OC Listeriaceae; Listeria.  
OX NCBI\_TaxID=1639;

RM (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN-EGD;  
RA Brehm K., Foerster S., Kretz J.;  
RT "The L. monocytogenes parA locus";  
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ010494; CAC06097.1;  
DR InterPro: IPR000707; ATPase\_PARA.  
DR InterPro: IPR000392; NitrogenaseII.  
DR Pfam: PF00991; Para; 1.  
DR PRINTS; PR00091; NITROGNASEII.  
SQ SEQUENCE 250 AA; 27186 MW; A77DA30B28C94CDD CRC64;

Query Match 100.0%; Score 26; DB 2; Length 250;  
Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5  
DB 228 LYDAK 232

## RESULT 12

Q8T3H1 PRELIMINARY; PRT; 250 AA.

ID Q8T3H1;  
AC Q8T3H1;  
DT 01-JUN-2002 (TREMBLrel. 21, Created)  
DT 01-JUN-2002 (TREMBLrel. 21, last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)  
DE Hypothetical 27.7 kDa protein.  
GN F55A12.2.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Pelodierinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;

RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Watson R.;  
RT "Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium.";  
RL Science 282:2012-2018(1998).  
RN (12)  
RP SEQUENCE FROM N.A.



RC STRAIN-BRISTOL N2;  
 RA Pauley A., Gattung S.;  
 RT "The sequence of *C. elegans* cosmid F55A12";  
 RL Submitted (Mar-1997) to the EMBL/GenBank/DBJ databases.  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Waterston R.;  
 RL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF003130; AM15579.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 250 AA; 27678 MW; BCD8E30A54E52215 CRC64;

Query Match 100.0%; Score 26; DB 5; Length 250;  
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5  
 Db 54 LYDAK 58

RESULT 13  
 Q926W7 PRELIMINARY; PRT; 253 AA.  
 AC Q926W7;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Partition protein, Para homolog.  
 GN PARA OR LMO2791 OR LIN2923.  
 OS Listeria monocytogenes, and  
 OS Listeria innocua.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
 OC Listeriaceae; Listeria.  
 OX NCBI\_TaxID=1639, 1642;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-L.monocytogenes, and L.innocua;  
 RC STRAIN-BD-E / SEROVAR 1/2A, AND CLIP 11262 / SEROVAR 6A;  
 RX MEDLINE-21537279; PubMed-11679669;  
 RA Glaser P., Frangoul L., Buchrieser C., Rusnlok C., Amend A.,  
 RA Baquero F., Berche P., Bioecker H., Brandt P., Chakraborty T.,  
 RA Charit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,  
 RA Dommann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,  
 RA Ertlen K.-D., Fslin H., Garcia-del Portillo F., Garrido P.,  
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
 RA Jones L.-M., Kaerst U., Kretz J., Kuhn M., Kunst F., Kurephat G.,  
 RA Medueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,  
 RA Nordiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,  
 RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,  
 RA Vazquez-Boland J.-A., Voss H., Weiland J., Cossart P.;  
 RT "Comparative genomics of *Listeria* species.";  
 RL Science 294:849-852(2001).  
 DR EMBL: AL591984; CAD01004.1; -  
 DR EMBL: AL596174; CAC98148.1; -  
 DR L1stLIst; LIN2923; -  
 DR L1stLIst; LMO2791; -  
 DR InterPro: IPR000707; ATPase\_Para.  
 DR Pfam: PF00991; Para; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 253 AA; 27470 MW; 70F307B95EFF33AC CRC64;

Query Match 100.0%; Score 26; DB 16; Length 253;  
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5  
 Db 231 LYDAK 235

RESULT 14

O9PJ15  
 ID O9PJ15 PRELIMINARY; PRT; 259 AA.  
 AC O9PJ15;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE Periplasmic protein.  
 GN C10111.  
 OS Campylobacter jejuni.  
 OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;  
 OC Campylobacter.  
 OX NCBI\_TaxID=197;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-NCTC 11168;  
 RX MEDLINE-20150912; PubMed-10688204;  
 RA Parkhill J., Wren B.W., Mungall K., Kettle J.M., Churcher C.,  
 RA Basham D., Chillingworth T., Davies R.M., Feltham S., Holtroyd S.,  
 RA Jagers K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,  
 RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,  
 RA Whitehead S., Barrell B.G.;  
 RT "The genome sequence of the food-borne pathogen *Campylobacter jejuni*  
 RT reveals hypervariable sequences.";  
 RL Nature 403:665-668(2000).  
 DR EMBL: AL139074; CAB72595.1; -  
 KW Complete proteome.  
 SQ SEQUENCE 259 AA; 29294 MW; AC4018A070B1D36F CRC64;

Query Match 100.0%; Score 26; DB 16; Length 259;  
 Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5  
 Db 218 LYDAK 222

RESULT 15  
 Q9P7L6 PRELIMINARY; PRT; 281 AA.  
 ID Q9P7L6;  
 AC Q9P7L6;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Putative methyltransferase.  
 GN SPBC21C3.07C.  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomyces.  
 OX NCBI\_TaxID=4896;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-972H-;  
 RA Seeger K., Harris D., Lyne M., Rajandream M.A., Barrell B.G.;  
 RL Submitted (Feb-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL157918; CAB76043.1; -  
 DR InterPro: IPR001601; Methyltransf.  
 DR InterPro: IPR000051; SAM\_bind.  
 KW Methyltransferase; Transferase.  
 SQ SEQUENCE 281 AA; 32732 MW; 222B60A6AA44B039 CRC64;

Query Match 100.0%; Score 26; DB 3; Length 281;  
 Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5  
 Db 164 LYDAK 168

Search completed: July 11, 2003, 11:59:05  
 Job time: 5.18295 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:42:16 ; Search time 2.86902 Seconds  
(without alignments)  
232.223 Million cell updates/sec

Title: US-10-091-442-8

Perfect score: 28  
Sequence: 1 NEMK 5

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A.Geneseq-101002:\*

- 1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*
- 2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*
- 3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*
- 4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*
- 5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*
- 6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*
- 7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*
- 8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*
- 9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*
- 10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*
- 11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*
- 12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*
- 13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*
- 14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*
- 15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*
- 16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*
- 17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*
- 18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*
- 19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*
- 20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*
- 21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*
- 22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*
- 23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28	100.0	5	15	AA857114
2	28	100.0	163	22	AA689325
3	28	100.0	380	15	AA848379
4	28	100.0	380	20	AAV08254
5	28	100.0	380	21	AA824142
6	28	100.0	380	22	AA83075
7	28	100.0	428	21	AA631168
8	28	100.0	434	20	AAV08215
9	28	100.0	441	21	AA631167
10	28	100.0	479	21	AA631166

11	28	100.0	847	20	AAV22237
12	28	100.0	2057	21	AA610667
13	25	89.3	32	22	AA661548
14	25	89.3	32	22	AA661550
15	25	89.3	32	22	AA661551
16	25	89.3	32	22	AA661552
17	25	89.3	32	22	AA661553
18	25	89.3	32	22	AA661556
19	25	89.3	37	22	AA661557
20	25	89.3	43	21	AA633870
21	25	89.3	52	23	AA608249
22	25	89.3	74	21	AA602988
23	25	89.3	90	21	AA603363
24	25	89.3	101	23	ABP34745
25	25	89.3	126	22	AA607397
26	25	89.3	161	23	ABP42805
27	25	89.3	172	21	AA608094
28	25	89.3	178	21	AA608093
29	25	89.3	192	21	AA608092
30	25	89.3	196	22	AA665978
31	25	89.3	214	22	AA670664
32	25	89.3	229	23	AA615440
33	25	89.3	288	23	ABP28086
34	25	89.3	299	21	AA605951
35	25	89.3	315	13	AA622214
36	25	89.3	332	13	AA622213
37	25	89.3	332	13	AA622218
38	25	89.3	337	22	AB600104
39	25	89.3	337	22	AB607578
40	25	89.3	392	22	AA681550
41	25	89.3	398	13	AA622212
42	25	89.3	415	13	AA622211
43	25	89.3	415	13	AA622217
44	25	89.3	419	15	AA63440
45	25	89.3	420	20	AAV08221

#### ALIGNMENTS

RESULT 1	AA857114	standard: peptide: 5 AA.	Human KDR signal t
AA857114	AA857114	standard: peptide: 5 AA.	L. mesenteroides a
XX	XX	XX	Peptide WINZIPA1 u
XX	XX	XX	Peptide WINZIPA3 u
XX	XX	XX	Peptide WINZIPA4 u
XX	XX	XX	Peptide WINZIPA5 u
XX	XX	XX	Peptide WINZIPA6 u
XX	XX	XX	Peptide WINZIPA9 u
XX	XX	XX	Peptide WINZIPA1
XX	XX	XX	Human secreted pro
XX	XX	XX	HA epitope tag. S
XX	XX	XX	Human secreted pro
XX	XX	XX	Human secreted pro
XX	XX	XX	Human ORF718 prot
XX	XX	XX	Human polypeptide
XX	XX	XX	Human ovarian anti
XX	XX	XX	Arabidopsis thalia
XX	XX	XX	Arabidopsis thalia
XX	XX	XX	Arabidopsis thalia
XX	XX	XX	Drosophila melanog
XX	XX	XX	S cerevisiae apopt
XX	XX	XX	Human drug metabol
XX	XX	XX	Streptococcus poly
XX	XX	XX	protein deduced fr
XX	XX	XX	Sequence of interl
XX	XX	XX	Sequence of interl
XX	XX	XX	Sequence of interl
XX	XX	XX	Novel human diagno
XX	XX	XX	Novel human diagno
XX	XX	XX	S. epidermidis ope
XX	XX	XX	Sequence of interl
XX	XX	XX	Sequence of interl
XX	XX	XX	Staphylococcus epi
XX	XX	XX	Staphylococcus hom

PT New megakaryocyte differentiation factor - isolated from human  
PT epidermoid carcinoma cells, used to treat conditions involving a  
XX decrease in platelets  
PS Claim 1: Page 20; 47pp; English.  
XX  
CC Human MDF (see AAR48379) can be isolated from a culture of human  
CC epidermoid carcinoma A431 cells in protein-free medium. The MDF  
CC stimulates differentiation of megakaryocytes from myeloid cells  
CC in the presence of IL-3. The MDF acts in vivo as a thrombopoietin  
CC making it useful for treatment of diseases involving a decrease  
CC in platelet number (esp. thrombocytopenia) such as occurs in bone  
CC marrow transplantation and in chemotherapy. MDF has mol.wt.  
CC 55-57kD by SDS-PAGE, exhibits an isoelectric point of 6.0-7.0 and  
CC contains an amino acid sequence comprising at least one of the  
CC sequences AAR57107-R57115.  
SQ Sequence 5 AA;  
Query Match 100.0%; Score 28; DB 15; Length 5;  
Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 NTEMK 5  
DB 1 NTEMK 5  
RESULT 2  
AAG89325  
ID AAG89325 standard; Protein; 163 AA.  
XX  
AC AAG89325;  
XX  
DT 11-SEP-2001 (first entry)  
XX  
DE Human secreted protein, SEQ ID NO: 445.  
XX  
KW Human; secreted protein; gene therapy; vaccine; treatment; diagnosis;  
KW GENSET.  
XX  
OS Homo sapiens.  
XX  
PN WO200142451-A2.  
PD 14-JUN-2001.  
XX  
PF 07-DEC-2000; 2000MO-IB01938.  
XX  
PR 08-DEC-1999; 99US-0169629.  
PR 06-MAR-2000; 2000US-0187470.  
XX  
PA (GENSET) GENSET.  
XX  
PI Dumas Milne Edwards J, Bougueleret L, Jobert S;  
PI WPI: 2001-367870/38.  
DR N-PSDB: AAR64928.  
XX  
PT Full length GENSET human nucleic acids encoding potentially secreted  
PT proteins, useful in gene therapy and vaccination against a variety of  
PT diseases, and for diagnosis of those diseases -  
XX  
PS Claim 21; Page 903; 921pp; English.  
XX  
CC The invention relates to full length GENSET human nucleic acids encoding  
CC potentially secreted proteins. The nucleic acids and the polypeptides  
CC they encode may be used in the prevention, treatment and diagnosis of  
CC diseases associated with inappropriate GENSET gene expression. For  
CC example, they be used to treat disorders associated with decreased  
CC GENSET gene expression by rectifying mutations or deletions in a  
CC patient's genome that affect the activity of GENSET or by supplementing  
CC the patients own production of GENSET polypeptides. Conversely,

CC antisense nucleic acid molecules may be administered to down regulate  
CC GENSET expression by binding with the cells' own genes and preventing  
CC their expression. The sense and antisense nucleic acids may also be  
CC used as DNA probes in diagnostic assays to detect and quantitate the  
CC presence of similar nucleic acid sequences in samples, and hence to  
CC determine which patients may be in need of restorative therapy.  
CC The GENSET polypeptides may be used as antigens in the production of  
CC antibodies and in assays to identify modulators (agonists and  
CC antagonists) of GENSET polypeptide expression and activity. The  
CC present sequence is a GENSET polypeptide of the invention.  
SQ Sequence 163 AA;  
Query Match 100.0%; Score 28; DB 22; Length 163;  
Best Local Similarity 100.0%; Pred. No. 63;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 NTEMK 5  
DB 159 NTEMK 163  
RESULT 3  
AAR48379  
ID AAR48379 standard; Protein; 380 AA.  
XX  
AC AAR48379;  
XX  
DT 16-AUG-1994 (first entry)  
XX  
DE Human megakaryocyte differentiation factor.  
XX  
DE Human megakaryocyte differentiation factor.  
XX  
KW Human megakaryocyte differentiation factor; MDF; thrombopoietin;  
KW haematopoietic stimulating factor; thrombocytopenia; platelet;  
KW bone marrow transplantation; cancer chemotherapy.  
XX  
OS Homo sapiens.  
XX  
PN EP583884-A.  
XX  
PD 23-FEB-1994.  
XX  
PE 19-JUL-1993; 93EP-0305654.  
XX  
PR 17-JUL-1992; 92JP-0212305.  
PR 04-MAR-1993; 93JP-0067339.  
XX  
PA (SUNR) SUNTORI LTD.  
PA (TSUJ/) TSUJIMOTO M.  
XX  
PI Ishida N, Iwasa F, Kurihara T, Miura K, Nakazato H;  
PI Tsujimoto M, Tsuruoka N, Yamaguchi N, Yamaichi K;  
XX  
DR WPI: 1994-058782/08.  
DR N-PSDB: AAO56670.  
XX  
PT New megakaryocyte differentiation factor - isolated from human  
PT epidermoid carcinoma cells, used to treat conditions involving a  
PT decrease in platelets  
XX  
PS Claim 7; Page 30-32; 47pp; English.  
XX  
CC Human MDF can be isolated from a culture of human epidermoid  
CC carcinoma A431 cells in protein-free medium. The MDF stimulates  
CC differentiation of megakaryocytes from myeloid cells in the presence  
CC of IL-3. The MDF acts in vivo as a thrombopoietin making it useful  
CC for treatment of diseases involving a decrease in platelet number  
CC (esp. thrombocytopenia) such as occurs in bone marrow  
CC transplantation and in chemotherapy. MDF has mol.wt. 55-57kD by  
CC SDS-PAGE and exhibits an isoelectric point of 6.0-7.0.  
XX  
SQ Sequence 380 AA;

Query Match 100.0%; Score 28; DB 15; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NTEMK 5  
 |||||  
 DB 284 NTEMK 288

## RESULT 4

AA08254  
 ID AA08254 standard; Protein; 380 AA.

AC AA08254;

DT 14-JUL-1999 (first entry)

DE Human megalin protein.

KW Megalin; mesangial cell; treatment; diagnosis; disease; Iga nephropathy;

KM human; rat; murine.

XX Homo sapiens.

PN WO915652-A1.

PD 01-APR-1999.

PF 22-SEP-1998; 98WO-JP04269.

PR 22-SEP-1997; 97JP-0275302.

PA (KURO/) KUROKAWA K.

PA (MIYA/) MIYATA T.

PI Miyata T;

DR WPI: 1999-276983/23.

DR N-PSDB; AAX56712.

PT Megalin protein expressed specifically in mesangial cells

PS Claim 1; Page 62-64; 100pp; Japanese.

CC This invention describes the isolation of novel megalin nucleic acid and

CC proteins from human, rat and mouse tissue. This protein is expressed

CC specifically in mesangial cells. The products of the invention are

CC useful for the treatment and diagnosis of diseases involving mesangial

CC cells, such as Iga nephropathy.

SO Sequence 380 AA;

OY 1 NTEMK 5  
 |||||  
 DB 284 NTEMK 288

## RESULT 5

AA08254  
 ID AA08254 standard; Protein; 380 AA.

AC AA08254;

DT 30-JAN-2001 (first entry)

DE Human megalin protein sequence SEQ ID NO:2.

KW Megalin; mesangium-predominant gene; serpin regulated; nephropathy;

KM Iga; immunoglobulin A; detection; renal function; renal disorder;

KM diagnosis; biological sample; blood; urine.

OS Homo sapiens.

PN WO200057189-A1.

PD 28-SEP-2000.

PF 17-MAR-2000; 2000WO-JP01646.

PR 19-MAR-1999; 99JP-0075305.

PR 28-OCT-1999; 99JP-0306623.

PA (KURO/) KUROKAWA K.

PA (FUSO) FUSO PHARM IND LTD.

PA (MIYA/) MIYATA T.

PI Miyata T;

DR WPI: 2000-611642/58.

DR N-PSDB; AAX99294.

PT Evaluating renal function comprises assaying megalin protein in

PT biological sample

PS Example 2; Page 66-69; 93pp; Japanese.

CC The present invention describes a method for evaluating renal function.

CC The method comprises assaying megalin protein in biological sample. Also

CC described are: (1) use of an anti-megalin protein antibody for diagnosing

CC renal function; and (2) a kit for detecting megalin protein comprising:

CC (a) anti-megalin protein antibody attached to solid magnetic particles;

CC (b) direct or indirect fixing for the antibody to the particles; and

CC (c) a magnet. The process is useful for evaluating renal function and

CC diagnosing renal disorders by assaying megalin protein in biological

CC samples (preferably urine or blood). The process is reproducible and

CC gives accurate results. The present sequence represents the human megalin

CC protein, which is given in the exemplification of the present invention.

SO Sequence 380 AA;

OY 1 NTEMK 5  
 |||||  
 DB 284 NTEMK 288

## RESULT 6

AA083075  
 ID AA083075 standard; Protein; 380 AA.

AC AA083075;

DT 10-JUL-2001 (first entry)

DE Human megalin protein.

KW Human; megalin; mesangial cell proliferative nephritis; nephrotropic;

KM transgenic mouse; glomerular disease; animal model; drug screening.

OS Homo sapiens.

PN WO200124628-A1.

PD 12-APR-2001.

PF 06-OCT-2000; 2000WO-JP06988.

PR 06-OCT-1999; 99JP-0285736.

PA (KURO/) KUROKAWA K.  
PA (MIYA/) MIYATA T.  
XX  
PI Miyata T;  
XX  
DR WPI: 2001-300136/31.  
DR N-PSDB: AAF82438.  
XX  
XX Mouse model for mesangial cell proliferative nephritis for development  
PT and screening of new treatments -  
XX  
PS Example 4; Page 44-46; 62pp; Japanese.  
XX  
CC The present sequence is human megin. The human megin coding  
CC sequence may be introduced into a mouse to produce an animal model of  
CC mesangial cell proliferative nephritis. The symptoms include  
CC enlargement of the mesangial base region, sedimentation of an immune  
CC complex and an increase in mesangial cells. The animal model is useful  
CC for analysing the pathology of chronic glomerular diseases and for  
CC screening compositions for prevention and treatment of the diseases.  
CC Highly uniform models can be made easily and in large numbers using  
CC this method.  
XX  
SQ Sequence 380 AA;  
  
Query Match 100.0%; Score 28; DB 22; Length 380;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 NEMK 5  
DB 284 NEMK 288  
  
RESULT 7  
AAG31168  
ID AAG31168 standard; Protein: 428 AA.  
XX  
AC AAG31168;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 37385.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0128845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.

```

PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 08-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.

```

```

PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 100.0%; Score 28; DB 21; Length 428;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYEMK 5
Db 393 NYEMK 397

RESULT 8
AAV08215
ID AAV08215 standard; Protein: 434 AA.
XX
AC AAV08215;
XX
DT 09-JUL-1999 (first entry)
XX
DE Staphylococcus haemolyticus FemA protein.
XX
XX FemA: identification; detection; therapy; infection; femB;
KM amplification; genotyping; gram-positive bacteria; vaccine.
XX
XX Staphylococcus haemolyticus.
XX
FH Key Location/Qualifiers
FT Misc-difference 421
FT Misc-difference /note- "in frame stop codon"
FT Misc-difference 425
FT Misc-difference /note- "in frame stop codon"
XX
PN WO9916780-A2.
XX
PD 08-APR-1999.
XX
XX 28-SEP-1998; 98WO-BE00141.
XX
XX 26-SEP-1997; 97EP-0870146.
XX
PA (BENA-) BELGIAN MIN NAT DEFENCE.
PI (UYLO-) UNIV CATHOLIQUE LOUVAIN.
XX
XX Gala J, Vannuffel P;
XX
DR WPI: 1999-287521/24.
DR N-PSDB: AAX37798.
XX
XX New Staphylococcus-specific oligonucleotides
XX
PS Claim 16; Fig 6a-b; 48pp; English.
XX
CC This invention describes novel Staphylococcus-specific oligonucleotides
CC based on the consensus femA nucleotide sequence which are used to
CC develop products for the identification, detection and therapy of
CC infections. The oligonucleotides can be used for the genetic
CC amplification, the identification and/or quantification of various femA

```

CC sequences which are specific to known or unknown Staphylococci species.  
CC Since the fema sequence is similar to the femB sequence, the  
CC oligonucleotides can also be used for the molecular genotyping of femB  
CC genes of different Staphylococci species or other gram-positive bacteria.  
CC The fema nucleic acid can also be used in therapeutic applications.  
CC They can also be used to identify inhibitors, e.g. antibodies or  
CC antisense oligonucleotides, for blocking expression of the fema  
CC nucleotide sequences. They can also be used for producing vaccines  
CC against Staphylococci infections.  
XX  
SQ Sequence 434 AA;  
  
Query Match 100.0%; Score 28; DB 20; Length 434;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 NTEMK 5  
DB 29 NTEMK 33  
  
RESULT 9  
AAG31167  
ID AAG31167 standard; Protein; 441 AA.  
XX  
AC AAG31167;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 37384.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 990S-0121825.  
PR 05-MAR-1999; 990S-0123180.  
PR 09-MAR-1999; 990S-0123548.  
PR 23-MAR-1999; 990S-0125788.  
PR 25-MAR-1999; 990S-0126264.  
PR 29-MAR-1999; 990S-0126785.  
PR 01-APR-1999; 990S-0127462.  
PR 06-APR-1999; 990S-0128234.  
PR 08-APR-1999; 990S-0128714.  
PR 16-APR-1999; 990S-0129845.  
PR 19-APR-1999; 990S-0130077.  
PR 21-APR-1999; 990S-0130449.  
PR 23-APR-1999; 990S-0130510.  
PR 28-APR-1999; 990S-0130891.  
PR 30-APR-1999; 990S-0132048.  
PR 30-APR-1999; 990S-0132048.  
PR 04-MAY-1999; 990S-0132484.  
PR 05-MAY-1999; 990S-0132485.  
PR 06-MAY-1999; 990S-0132486.  
PR 07-MAY-1999; 990S-0132487.  
PR 11-MAY-1999; 990S-0132863.  
PR 14-MAY-1999; 990S-0134218.  
PR 14-MAY-1999; 990S-0134219.  
PR 14-MAY-1999; 990S-0134221.  
PR 14-MAY-1999; 990S-0134370.  
PR 18-MAY-1999; 990S-0134768.  
PR 19-MAY-1999; 990S-0134941.  
PR 20-MAY-1999; 990S-0135124.

PR 21-MAY-1999; 990S-0135353.  
PR 24-MAY-1999; 990S-0135629.  
PR 25-MAY-1999; 990S-0136021.  
PR 27-MAY-1999; 990S-0136392.  
PR 28-MAY-1999; 990S-0136782.  
PR 01-JUN-1999; 990S-0137222.  
PR 03-JUN-1999; 990S-0137528.  
PR 04-JUN-1999; 990S-0137502.  
PR 07-JUN-1999; 990S-0137724.  
PR 08-JUN-1999; 990S-0138094.  
PR 10-JUN-1999; 990S-0138540.  
PR 10-JUN-1999; 990S-0138847.  
PR 14-JUN-1999; 990S-0139119.  
PR 16-JUN-1999; 990S-0139452.  
PR 16-JUN-1999; 990S-0139453.  
PR 17-JUN-1999; 990S-0139492.  
PR 18-JUN-1999; 990S-0139454.  
PR 18-JUN-1999; 990S-0139455.  
PR 18-JUN-1999; 990S-0139456.  
PR 18-JUN-1999; 990S-0139457.  
PR 18-JUN-1999; 990S-0139459.  
PR 18-JUN-1999; 990S-0139460.  
PR 18-JUN-1999; 990S-0139461.  
PR 18-JUN-1999; 990S-0139462.  
PR 18-JUN-1999; 990S-0139463.  
PR 18-JUN-1999; 990S-0139750.  
PR 18-JUN-1999; 990S-0139763.  
PR 21-JUN-1999; 990S-0139817.  
PR 22-JUN-1999; 990S-0139899.  
PR 23-JUN-1999; 990S-0140353.  
PR 23-JUN-1999; 990S-0140354.  
PR 24-JUN-1999; 990S-0140695.  
PR 28-JUN-1999; 990S-0140823.  
PR 29-JUN-1999; 990S-0140991.  
PR 30-JUN-1999; 990S-0141287.  
PR 01-JUL-1999; 990S-0141842.  
PR 01-JUL-1999; 990S-0142154.  
PR 02-JUL-1999; 990S-0142055.  
PR 06-JUL-1999; 990S-0142390.  
PR 08-JUL-1999; 990S-0142803.  
PR 09-JUL-1999; 990S-0142820.  
PR 12-JUL-1999; 990S-0142977.  
PR 13-JUL-1999; 990S-0143542.  
PR 14-JUL-1999; 990S-0143624.  
PR 15-JUL-1999; 990S-0144005.  
PR 16-JUL-1999; 990S-0144085.  
PR 16-JUL-1999; 990S-0144086.  
PR 19-JUL-1999; 990S-0144325.  
PR 19-JUL-1999; 990S-0144331.  
PR 19-JUL-1999; 990S-0144332.  
PR 19-JUL-1999; 990S-0144333.  
PR 19-JUL-1999; 990S-0144334.  
PR 19-JUL-1999; 990S-0144335.  
PR 20-JUL-1999; 990S-0144352.  
PR 20-JUL-1999; 990S-0144632.  
PR 20-JUL-1999; 990S-0144884.  
PR 21-JUL-1999; 990S-0144814.  
PR 21-JUL-1999; 990S-0145086.  
PR 21-JUL-1999; 990S-0145088.  
PR 22-JUL-1999; 990S-0145085.  
PR 22-JUL-1999; 990S-0145087.  
PR 22-JUL-1999; 990S-0145089.  
PR 22-JUL-1999; 990S-0145192.  
PR 23-JUL-1999; 990S-0145145.  
PR 23-JUL-1999; 990S-0145218.  
PR 23-JUL-1999; 990S-0145224.  
PR 26-JUL-1999; 990S-0145276.  
PR 27-JUL-1999; 990S-0145913.  
PR 27-JUL-1999; 990S-0145918.  
PR 27-JUL-1999; 990S-0145919.  
PR 28-JUL-1999; 990S-0145951.  
PR 02-AUG-1999; 990S-0146386.



PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 23-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.

PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
  
Query Match 100.0%; Score 28; DB 21; Length 441;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 NYEMK 5  
Db 406 NYEMK 410  
  
RESULT 10  
AAG31166  
ID AAG31166 standard; Protein; 479 AA.  
XX  
AC AAG31166;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 37383.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 09-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.

PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148365.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159684.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 100.0%; Score 28; DB 21; Length 479;  
Best Local Similarity 100.0%; Pred. No. 2e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NYEMK 5  
11111  
Db 444 NYEMK 448

## RESULT 11

AAV22237  
ID AAV22237 standard; Protein: 847 AA.

AC AAV22237;

DT 20-SEP-1999 (first entry)

DE Human KDR signal transduction inducer protein sequence.

XX KDR signal transduction inducer protein; human; diabetic retinopathy;  
KM vascular endothelial cell growth receptor; abnormal neovascularisation;  
KM kinase insert domain-containing receptor; solid tumour proliferation;  
KM gene therapy; metastasis; chronic rheumatoid arthritis; psoriasis;  
KM retinopathy; retinopathy of prematurity.

XX Homo sapiens.

OS MO9931238-A1.

PN 24-JUN-1999.

PF 11-DEC-1998; 98MO-JP05612.

PR 12-DEC-1997; 97JP-0343474.

PA (KYOW) KYOWA HAKKO KOGYO KK.

PA (SHIB/) SHIBUYA M.

PI Shibuya M, Yabana N;

XX WPI: 1999-405033/34.

DR N-PSDB: AAX84667.

PT KDR signal transduction inducing protein and antibodies to it

PS Claim 1; Page 68-74; 82pp; Japanese.

XX This sequence is the protein of the invention, which induces signal  
CC transduction of the vascular endothelial cell growth receptor KDR (kinase  
CC insert domain-containing receptor) by binding to its intracellular  
CC domain. The protein can be used in the investigation, diagnosis and  
CC treatment (including gene therapy) of diseases in which abnormal  
CC neovascularisation takes place, such as solid tumour proliferation,  
CC metastasis, chronic rheumatoid arthritis, psoriasis and retinopathy  
CC (including diabetic retinopathy and retinopathy of prematurity). It may  
CC be used as a screen for candidate KDR signal transduction inhibitors for  
CC therapeutic use.

SO Sequence 847 AA;

Query Match 100.0%; Score 28; DB 20; Length 847;  
Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NYEMK 5  
11111  
Db 438 NYEMK 442

## RESULT 12

AAAB10667  
ID AAB10667 standard; Protein: 2057 AA.

AC AAB10667;

XX

DT 19-JAN-2001 (first entry)

XX L. mesenteroides alternan sucrose protein.

DE Alternan sucrose; glucosyltransferase; fructose; cosmetic; foodstuff;  
KM syrup.

XX Leuconostoc mesenteroides.

XX DEL9905069-A1.

PN 10-AUG-2000.

PD 08-FEB-1999; 99DE-1005069.

PE 08-FEB-1999; 99DE-1005069.

PR (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.

PA (PLAC) MAX PLANCK GES FORDERUNG WISSENSCHAFTEN.

XX Kossmann J, Welsh T, Quanz M, Knuth K;

PI WPI: 2000-550294/51.

DR N-PSDB: AAA97904.

PT New nucleic acid encoding recombinant Leuconostoc mesenteroides  
PT alternan sucrose protein and methods of alternan and fructose  
PT production

PS Claim 1a; Page 30-36; 64pp; German.

XX This invention describes a novel nucleic acid molecule (I) encoding an  
CC alternan sucrose (E.C. 2.4.1.140 - an enzyme, that belongs to the  
CC glucosyltransferase group) The recombinant, purified alternan sucrose  
CC gene is useful for the fermentative production of alternan (a  
CC carbohydrate) and/or fructose by secreting the enzyme into a  
CC saccharose-containing culture medium. Alternatively, the enzyme is  
CC contacted with a saccharose-containing solution. The alternan and/or  
CC fructose is then isolated from the medium. Cosmetic products or  
CC foodstuffs containing alternan can be produced. Recombinant production of  
CC alternan sucrose is advantageous as it provides a cost effective means of  
CC producing fructose for high fructose containing syrups, production of  
CC which previously has been achieved by costly production from maize  
CC starch. This sequence represents the Leuconostoc mesenteroides alternan  
CC sucrose protein which is described in the method of the invention.

SO Sequence 2057 AA;

Query Match 100.0%; Score 28; DB 21; Length 2057;  
Best Local Similarity 100.0%; Pred. No. 9.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NYEMK 5  
11111  
Db 1003 NYEMK 1007

## RESULT 13

AAAB61548  
ID AAB61548 standard; Peptide: 32 AA.

AC AAB61548;

XX 03-APR-2001 (first entry)

DE Peptide WINZIPAL used to identify hetero-associating peptides.

XX Hetero-associating coiled-coil peptide; heterodimerisation.

XX Unidentified.

OS WO200100814-A2.

XX

PD	04-JAN-2001.
XX	
PF	26-JUN-2000; 2000WO-EP05922.
XX	
PR	25-JUN-1999; 99US-0344096.
XX	
PA	(UYZU-) UNIV ZUERICH.
XX	
PI	Plueckthun A, Arndt K, Mueller K, Pelletier J;
XX	
DR	WPI: 2001-137954/14.
XX	
PT	Identifying heteroassociating (poly)peptides involves designing an
PT	appropriate coiled-coil library and screening by using a library versus
PT	library approach
XX	
PS	Claim 5; Page 40; 56pp; English.
XX	
CC	The present invention relates to a method for identifying
CC	hetero-associating coiled-coil peptides. The method comprises providing a
CC	library of peptides with the general formula of AAB61546, and a second
CC	library of peptides with the general formula of AAB61547.
CC	Hetero-association of peptides from the two libraries can then be
CC	screened for via a screenable or selectable property caused by the
CC	hetero-association of the two peptides. The identified peptides can be
CC	used for heterodimerisation of fusion proteins. The present sequence is a
CC	hetero-associating peptide that can be used in method of the present
CC	invention.
XX	
SQ	Sequence 32 AA;
XX	
Query Match	89.3%; Score 25; DB 22; Length 32;
Best Local Similarity	80.0%; Pred. No. 55;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0	
QY	1 NYEMR 5
	:
Db	15 NYELK 19
XX	
RESULT 14	
AAB61550	
ID	AAB61550 standard; Peptide: 32 AA.
XX	
AC	AAB61550;
XX	
DT	03-APR-2001 (first entry)
XX	
DE	Peptide W1NZIP3 used to identify hetero-associating peptides.
XX	
KW	Hetero-associating coiled-coil peptide; heterodimerisation.
XX	
OS	Unidentified.
XX	
NN	WO200100814-A2.
XX	
PD	04-JAN-2001.
XX	
PF	26-JUN-2000; 2000WO-EP05922.
XX	
PR	25-JUN-1999; 99US-0344096.
XX	
PA	(UYZU-) UNIV ZUERICH.
XX	
PI	Plueckthun A, Arndt K, Mueller K, Pelletier J;
XX	
DR	WPI: 2001-137954/14.
XX	
PT	Identifying heteroassociating (poly)peptides involves designing an
PT	appropriate coiled-coil library and screening by using a library versus
PT	library approach
XX	
XS	Claim 5; Page 40; 56pp; English.

Query Match	Best Local Similarity	Score 25;	DB 22;	Length 32;
Matches 4; Conservative 1;	Mismatches 0;	Indels 0;	Gaps 0;	
1 NYEMK 5	:			
15 NYELK 19				
Sequence 32 AA;				
Query Match	Best Local Similarity	Score 25;	DB 22;	Length 32;
Matches 4; Conservative 1;	Mismatches 0;	Indels 0;	Gaps 0;	
1 NYEMK 5	:			
15 NYELK 19				
Sequence 32 AA;				
Query Match	Best Local Similarity	Score 25;	DB 22;	Length 32;
Matches 4; Conservative 1;	Mismatches 0;	Indels 0;	Gaps 0;	
1 NYEMK 5	:			
15 NYELK 19				
Sequence 32 AA;				

OY 1 NYEMK 5  
111:1  
Db 15 NYELK 19

Search completed: July 11, 2003, 11:54:25  
Job time : 3.86902 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:49:36 ; Search time 0.893971 Seconds  
(without alignments)  
164.563 Million cell updates/sec

Title: US-10-091-442-8

Perfect score: 28

Sequence: 1 NYEMK 5

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28	100.0	5	2	US-08-472-659-8
2	28	100.0	5	2	US-08-474-661-8
3	28	100.0	5	2	US-08-611-977-8
4	28	100.0	380	2	US-08-472-659-34
5	28	100.0	380	2	US-08-474-661-34
6	28	100.0	380	2	US-08-611-977-34
7	25	89.3	13	1	US-08-306-231-6
8	25	89.3	13	1	US-08-355-888A-22
9	25	89.3	13	3	US-08-693-697-22
10	25	89.3	99	3	US-08-946-329A-89
11	25	89.3	235	1	US-08-928-443-4
12	25	89.3	235	3	US-09-129-055-4
13	25	89.3	315	1	US-07-757-390-8
14	25	89.3	315	1	US-08-442-282-8
15	25	89.3	315	1	US-08-442-282-8
16	25	89.3	315	1	US-08-442-281-8
17	25	89.3	315	2	US-08-939-727-8
18	25	89.3	332	1	US-07-757-390-7
19	25	89.3	332	1	US-08-442-282-7
20	25	89.3	332	1	US-08-442-281-7
21	25	89.3	332	2	US-08-939-727-7
22	25	89.3	335	1	US-07-947-130-3
23	25	89.3	335	1	US-08-421-822-3
24	25	89.3	335	1	US-08-421-823-3
25	25	89.3	338	1	US-07-757-390-6
26	25	89.3	338	1	US-08-442-282-6
27	25	89.3	398	1	US-08-442-281-6

28	25	89.3	398	2	US-08-939-727-6	Sequence 6, Appl1
29	25	89.3	415	1	US-07-757-390-5	Sequence 5, Appl1
30	25	89.3	415	1	US-08-442-282-5	Sequence 5, Appl1
31	25	89.3	415	1	US-08-442-281-5	Sequence 5, Appl1
32	25	89.3	415	2	US-08-939-727-5	Sequence 2, Appl1
33	25	89.3	419	1	US-08-330-154-2	Sequence 5230, Ap
34	25	89.3	422	4	US-09-134-001C-5230	Sequence 4, Appl1
35	25	89.3	429	3	US-08-307-896-4	Sequence 4, Appl1
36	25	89.3	429	5	PCT-US95-11808-4	Sequence 2, Appl1
37	25	89.3	452	2	US-08-416-870C-2	Sequence 4, Appl1
38	25	89.3	534	4	US-09-321-276-4	Sequence 2, Appl1
39	25	89.3	554	4	US-08-916-481-2	Sequence 3, Appl1
40	25	89.3	563	4	US-08-916-481-3	Sequence 3, Appl1
41	25	89.3	1090	3	US-08-307-896-3	Sequence 3, Appl1
42	25	89.3	1090	3	US-08-726-214-4	Sequence 4, Appl1
43	25	89.3	1090	5	PCT-US95-11808-3	Sequence 3, Appl1
44	25	89.3	2254	2	US-08-577-010-3	Sequence 3, Appl1
45	25	89.3	2254	2	US-08-790-519-3	Sequence 3, Appl1

#### ALIGNMENTS

RESULT 1  
US-08-472-659-8  
Sequence 8, Application US/08472659  
Patent No. 5831030  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSUROUOKA, No. 5831030uo  
APPLICANT: NAKAZONO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. 5831030uhitro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAGUCHI, Kozo  
APPLICANT: YAMAGUCHI, No. 5831030omi  
TITLE OF INVENTION: MECAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/472,659  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 536  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-248  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:

LENGTH: 5 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-472-659-8

Query Match 100.0%; Score 28; DB 2; Length 5;  
Best Local Similarity 100.0%; Pred. No. 1.9e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYEMK 5  
|||||  
DB 1 NYEMK 5

## RESULT 2

US-08-474-661-8  
Sequence 8, Application US/08474661  
Patent No. 5874253  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSURUOKA, No. 5874253uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MURA, Kenju  
APPLICANT: ISHIDA, No. 5874253uhltro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, No. 5874253oml  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/474,661  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: REA, TERESA STANER  
REGISTRATION NUMBER: 30,427  
REFERENCE/DOCKET NUMBER: 001560-204  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-6620  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-474-661-8

Query Match 100.0%; Score 28; DB 2; Length 5;

Best Local Similarity 100.0%; Pred. No. 1.9e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYEMK 5  
|||||  
DB 1 NYEMK 5

## RESULT 3

US-08-611-977-8  
Sequence 8, Application US/08611977  
Patent No. 5972886  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSURUOKA, No. 5972886uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MURA, Kenju  
APPLICANT: ISHIDA, No. 5972886uhltro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, No. 5972886oml  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/611,977  
FILING DATE: 06-MAR-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-204  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-611-977-8

Query Match 100.0%; Score 28; DB 2; Length 5;  
Best Local Similarity 100.0%; Pred. No. 1.9e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYEMK 5  
|||||  
DB 1 NYEMK 5



RESULT 4  
US-08-472-659-34  
Sequence 34, Application US/08472659  
Patent No. 5831030  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSURUOKA, No. 5831030uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. 5831030uh1ro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, No. 5831030om1  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/472,659  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 536  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: MCGOWAN, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-248  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-472-659-34

Query Match 100.0%; Score 28; DB 2; Length 380;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYEMK 5  
DB 284 NYEMK 288

RESULT 5  
US-08-474-661-34  
Sequence 34, Application US/08474661  
Patent No. 5874253  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki

APPLICANT: TSURUOKA, No. 5874253uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. 5874253uh1ro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, No. 5874253om1  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/474,661  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: REA, TERESA STANEK  
REGISTRATION NUMBER: 30,427  
REFERENCE/DOCKET NUMBER: 001560-204  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-6620  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-474-661-34

Query Match 100.0%; Score 28; DB 2; Length 380;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYEMK 5  
DB 284 NYEMK 288

RESULT 6  
US-08-611-977-34  
Sequence 34, Application US/08611977  
Patent No. 5972886  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSURUOKA, No. 5972886uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. 5972886uh1ro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, No. 5972886om1  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34

;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Burns, Doane, Swecker & Mathis  
;; STREET: P.O. Box 1404  
;; CITY: Alexandria  
;; STATE: Virginia  
;; COUNTRY: United States  
;; ZIP: 22131-1404  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;;  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/611,977  
;; FILING DATE: 06-MAR-1996  
;; CLASSIFICATION: 435  
;;  
;; PRIORITY APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/091,028  
;; FILING DATE: 14-JUL-1993  
;;  
;; PRIORITY APPLICATION DATA:  
;; APPLICATION NUMBER: JP 4-212305  
;; FILING DATE: 17-JUL-1992  
;;  
;; PRIORITY APPLICATION DATA:  
;; APPLICATION NUMBER: JP 6-067339  
;; FILING DATE: 04-MAR-1993  
;;  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: McGowan, Malcolm K.  
;; REGISTRATION NUMBER: 39,300  
;; REFERENCE/DOCKET NUMBER: 001560-204  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (703) 836-6620  
;; TELEFAX: (703) 836-2051  
;;  
;; INFORMATION FOR SEQ ID NO: 34:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 380 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;;  
;; US-08-611-977-34

Query Match 100.0%; Score 28; DB 2; Length 380;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYEMK 5  
|||||  
Db 284 NYEMK 288

RESULT 7  
US-08-306-231-6  
;; Sequence 6, Application US/08306231  
;; Patent No. 5643748  
;;  
;; GENERAL INFORMATION:  
;; APPLICANT: Snodgrass, H. R.  
;; APPLICANT: Ciofili, Joseph  
;; APPLICANT: Zupancic, Thomas J.  
;; APPLICANT: Shafer, Alan W.  
;; TITLE OF INVENTION: Hu-B1,219, A NOVEL HEMATOPOIETIN  
;; TITLE OF INVENTION: RECEPTOR  
;; NUMBER OF SEQUENCES: 15  
;;  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Pennie & Edmonds  
;; STREET: 1155 Avenue of the Americas  
;; CITY: New York  
;; STATE: New York  
;; COUNTRY: USA  
;; ZIP: 10036-2711  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;;  
;; US-08-306-231-6

;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/306,231  
;; FILING DATE: 14-SEP-1994  
;; CLASSIFICATION: 435  
;;  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Poissant, Brian M.  
;; REGISTRATION NUMBER: 28,462  
;; REFERENCE/DOCKET NUMBER: 7225-076  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (212) 790-9090  
;; TELEFAX: (212) 869-9741/8864  
;;  
;; TELEEX: 66141 PENNIE  
;;  
;; INFORMATION FOR SEQ ID NO: 6:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 13 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS:  
;; TOPOLOGY: unknown  
;; MOLECULE TYPE: peptide  
;;  
;; US-08-306-231-6

Query Match 89.3%; Score 25; DB 1; Length 13;  
Best Local Similarity 80.0%; Pred. No. 3.9;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYEMK 5  
|||||  
Db 5 NYEMK 9

RESULT 8  
US-08-355-888A-22  
;; Sequence 22, Application US/08355888A  
;; Patent No. 5763211  
;;  
;; GENERAL INFORMATION:  
;; APPLICANT: Snodgrass, H. R.  
;; APPLICANT: Ciofili, Joseph  
;; APPLICANT: Zupancic, Thomas J.  
;; APPLICANT: Shafer, Alan W.  
;; TITLE OF INVENTION: Hu-B1,219, A NOVEL HUMAN HEMATOPOIETIN  
;; TITLE OF INVENTION: RECEPTOR  
;; NUMBER OF SEQUENCES: 31  
;;  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Pennie & Edmonds  
;; STREET: 1155 Avenue of the Americas  
;; CITY: New York  
;; STATE: New York  
;; COUNTRY: USA  
;; ZIP: 10036-2711  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;;  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/355,888A  
;; FILING DATE: 14-DEC-1994  
;; CLASSIFICATION: 435  
;;  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Poissant, Brian M.  
;; REGISTRATION NUMBER: 28,462  
;; REFERENCE/DOCKET NUMBER: 7225-078  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (212) 790-9090  
;; TELEFAX: (212) 869-9741/8864  
;;  
;; TELEEX: 66141 PENNIE  
;;  
;; INFORMATION FOR SEQ ID NO: 22:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 13 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS:  
;; TOPOLOGY: unknown  
;; MOLECULE TYPE: peptide  
;;  
;; US-08-355-888A-22

US-08-355-888A-22

Query Match 89.3%; Score 25; DB 1; Length 13;  
Best Local Similarity 80.0%; Pred. No. 3.9;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NYEMK 5  
111:1  
Db 5 NYELK 9

RESULT 9

US-08-693-697-22  
; Sequence 22, Application US/08693697  
; Patent No. 5869610  
; GENERAL INFORMATION:  
; APPLICANT: Snodgrass, H. R.  
; APPLICANT: Cioffi, Joseph  
; APPLICANT: Zupancic, Thomas J.  
; APPLICANT: Shafer, Alan W.  
; TITLE OF INVENTION: Hu-B1.219, A NOVEL HUMAN HEMATOPOIETIN  
; TITLE OF INVENTION: RECEPTOR  
; NUMBER OF SEQUENCES: 38  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0b  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/693,697  
; FILING DATE: 05-AUG-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Poissant, Brian M.  
; REGISTRATION NUMBER: 28,462  
; REFERENCE/DOCKET NUMBER: 8907-0037-999  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-493-4935  
; TELEFAX: 650-493-5556  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 13 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
US-08-693-697-22

Query Match 89.3%; Score 25; DB 2; Length 13;  
Best Local Similarity 80.0%; Pred. No. 3.9;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NYEMK 5  
111:1  
Db 5 NYELK 9

RESULT 10  
US-08-693-696-22  
; Sequence 22, Application US/08693696  
; Patent No. 6005080  
; GENERAL INFORMATION:  
; APPLICANT: Snodgrass, H. R.  
; APPLICANT: Cioffi, Joseph  
; APPLICANT: Zupancic, Thomas J.

APPLICANT: Shafer, Alan W.  
; TITLE OF INVENTION: Hu-B1.219, A NOVEL HUMAN HEMATOPOIETIN  
; TITLE OF INVENTION: RECEPTOR  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/693,696  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/355,888  
; FILING DATE: 14-DEC-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Poissant, Brian M.  
; REGISTRATION NUMBER: 28,462  
; REFERENCE/DOCKET NUMBER: 7225-078  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-9741/8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 13 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
US-08-693-696-22

Query Match 89.3%; Score 25; DB 3; Length 13;  
Best Local Similarity 80.0%; Pred. No. 3.9;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NYEMK 5  
111:1  
Db 5 NYELK 9

RESULT 11  
US-08-946-329A-89  
; Sequence 89, Application US/08946329A  
; Patent No. 6057091  
; GENERAL INFORMATION:  
; APPLICANT: Beachy, Philip A.  
; APPLICANT: Porter, Jeffrey A.  
; TITLE OF INVENTION: NOVEL HEDGEHOG-DERIVED POLYPEPTIDES  
; NUMBER OF SEQUENCES: 109  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 4225 Executive Square, Suite 1400  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; OPERATING SYSTEM: Windows 95  
; SOFTWARE: FASTSEQ for Windows Version 2.0b  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/946,329A  
; FILING DATE: 07-OCT-1997

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/061,323  
FILING DATE: 07-OCT-1996  
APPLICATION NUMBER: 08/729,743  
FILING DATE: 10-JUL-1996  
APPLICATION NUMBER: 08/567,357  
FILING DATE: 04-DEC-1995  
APPLICATION NUMBER: 08/349,498  
FILING DATE: 02-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Halle, Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 07265/140001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619/678-5070  
TELEFAX: 619/678-5099  
INFORMATION FOR SEQ ID NO: 89:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 99 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-946-329A-89

Query Match 89.3%; Score 25; DB 3; Length 99;  
Best Local Similarity 80.0%; Pred. No. 37;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 NYEMK 5  
Db 26 NYELK 30

RESULT 12  
US-08-928-443-4  
Sequence 4, Application US/08928443  
Patent No. 5795724  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Tang, Y. Tom  
APPLICANT: Corley, Neil C.  
TITLE OF INVENTION: HUMAN N-ACETYL TRANSFERASE  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/928,443  
FILING DATE: HERewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0361 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 235 amino acids

TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 517485  
US-08-928-443-4

Query Match 89.3%; Score 25; DB 1; Length 235;  
Best Local Similarity 80.0%; Pred. No. 96;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYEMK 5  
Db 25 NYQMK 29

RESULT 13  
US-09-129-055-4  
Sequence 4, Application US/09129055  
Patent No. 6017744

GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Tang, Y. Tom  
APPLICANT: Corley, Neil C.  
TITLE OF INVENTION: HUMAN N-ACETYL TRANSFERASE  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/129,055  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/928,443  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0361 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 235 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 517485  
US-09-129-055-4

Query Match 89.3%; Score 25; DB 3; Length 235;  
Best Local Similarity 80.0%; Pred. No. 96;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYEMK 5  
Db 25 NYQMK 29

RESULT 14  
US-07-757-390-8  
Sequence 8, Application US/07757390  
Patent No. 5453491  
GENERAL INFORMATION:  
APPLICANT: Takatsu, Kiyoshi  
APPLICANT: Tomioka, Akira  
APPLICANT: Takagi, Satoshi  
APPLICANT: Murata, Yoshiyuki  
TITLE OF INVENTION: Human And Murine Interleukin-5 Receptor  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/757,390  
FILING DATE: 19910910  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7005-030  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 790-9090  
TELEFAX: 212 8698864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 315 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: unknown  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-07-757-390-8  
Query Match 89.3%; Score 25; DB 1; Length 315;  
Best Local Similarity 80.0%; Pred. No. 1.3e+02;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 NYEMK 5  
DB 251 NYELK 255  
RESULT 15  
US-08-442-282-8  
Sequence 8, Application US/08442282  
Patent No. 5760204  
GENERAL INFORMATION:  
APPLICANT: Takatsu, Kiyoshi  
APPLICANT: Tomioka, Akira  
APPLICANT: Takagi, Satoshi  
APPLICANT: Murata, Yoshiyuki  
TITLE OF INVENTION: Human And Murine Interleukin-5 Receptor  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/442,282  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/757,390  
FILING DATE: 10-SEP-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7005-030  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 790-9090  
TELEFAX: 212 8698864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 315 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-442-282-8  
Query Match 89.3%; Score 25; DB 1; Length 315;  
Best Local Similarity 80.0%; Pred. No. 1.3e+02;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 NYEMK 5  
DB 251 NYELK 255  
Search completed: July 11, 2003, 12:02:11  
Job time: 1.89397 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:59:17 ; Search time 1.4657 Seconds  
(without alignments)  
397.148 Million cell updates/sec

Title: US-10-091-442-8

Perfect score: 28

Sequence: 1 NEMK 5

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 445758 seqs, 116419773 residues

Total number of hits satisfying chosen parameters: 445758

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications\_AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No	Score	Query Match	Length DB	ID	Description
1	28	100.0	5	9	US-10-091-442-8
2	28	100.0	5	10	US-09-140-719-8
3	28	100.0	163	10	US-09-731-872-445
4	28	100.0	380	9	US-10-091-442-34
5	28	100.0	380	10	US-09-140-719-34
6	28	100.0	399	9	US-10-029-180-64
7	28	100.0	584	9	US-09-995-749A-12
8	28	100.0	847	9	US-10-177-293-498
9	25	89.3	299	10	US-09-861-451A-16
10	25	89.3	446	10	US-09-853-386-69
11	25	89.3	554	9	US-09-878-672-2
12	25	89.3	554	10	US-09-800-336-4
13	25	89.3	563	9	US-09-878-672-3
14	25	89.3	625	10	US-09-853-386-63
15	25	89.3	626	10	US-09-853-386-64
16	25	89.3	626	10	US-09-853-386-65
17	25	89.3	626	10	US-09-853-386-96
18	25	89.3	643	10	US-09-853-386-70
19	24	85.7	42	10	US-09-864-761-42476

20	24	85.7	376	10	US-09-815-242-5669	Sequence 5669, App
21	24	85.7	377	10	US-09-815-242-12700	Sequence 12700, A
22	24	85.7	377	10	US-09-815-242-12726	Sequence 12726, A
23	24	85.7	557	9	US-10-081-872-206	Sequence 206, App
24	24	85.7	635	10	US-09-841-132-520	Sequence 520, App
25	24	85.7	896	9	US-10-210-296-5	Sequence 5, Appl1
26	24	85.7	898	9	US-10-270-875-37	Sequence 37, Appl1
27	24	85.7	898	9	US-10-270-878-37	Sequence 37, Appl1
28	24	85.7	898	9	US-10-270-786-37	Sequence 37, Appl1
29	24	85.7	898	9	US-10-270-710-37	Sequence 37, Appl1
30	24	85.7	898	9	US-10-270-859-37	Sequence 37, Appl1
31	23	82.1	5	10	US-09-832-312-67	Sequence 67, Appl1
32	23	82.1	67	10	US-09-864-761-33978	Sequence 33978, A
33	23	82.1	76	9	US-10-156-761-15015	Sequence 15015, A
34	23	82.1	83	9	US-10-091-572-237	Sequence 237, App
35	23	82.1	83	9	US-09-764-891-3348	Sequence 3348, App
36	23	82.1	110	10	US-09-815-242-13253	Sequence 13253, A
37	23	82.1	122	10	US-09-867-550-1862	Sequence 1862, App
38	23	82.1	132	9	US-10-093-766-44	Sequence 44, Appl1
39	23	82.1	132	9	US-10-171-311-89	Sequence 89, Appl1
40	23	82.1	132	9	US-10-197-666A-70	Sequence 70, Appl1
41	23	82.1	132	9	US-10-197-666A-146	Sequence 146, App
42	23	82.1	132	9	US-10-197-666A-148	Sequence 148, App
43	23	82.1	133	9	US-09-854-133-124	Sequence 124, App
44	23	82.1	133	9	US-10-197-666A-72	Sequence 72, Appl1
45	23	82.1	133	9	US-10-197-666A-74	Sequence 74, Appl1

#### ALIGNMENTS

RESULT 1  
US-10-091-442-8  
Sequence 8, Appl1  
Patent No. US20020164711A1  
Application US/10091442

GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
IMASA, Fuyuki  
TSURUOKA, No. US20020164711A1uo  
NAKAZATO, Hiroshi  
MIURA, Kenju  
ISHIDA, No. US20020164711A1uhlro  
KURIHARA, Tatsuya  
YAMACHIKI, Kozo  
YAMAGUCHI, No. US20020164711A1omi

TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/091,442  
FILING DATE: 07-Mar-2002  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/140, 719  
FILING DATE: 08-AUG-1998  
APPLICATION NUMBER: US 08/474, 661  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
APPLICATION NUMBER: JP 6-067339

FILED DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-2021  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 8:  
US-10-091-442-8

Query Match 100.0%; Score 28; DB 9; Length 5;  
Best Local Similarity 100.0%; Pred. No. 4e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYEMK 5  
DB 1 NYEMK 5

RESULT 2  
US-09-140-719-8  
Sequence 8, Application US/09140719  
Patent No. US20010026931A1  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IMASA, Fuyuki  
APPLICANT: TSUBOKAWA, No. US20010026931A1uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. US20010026931A1uh1ro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, No. US20010026931A1om1  
TITLE OF INVENTION: MEKANARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.

REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-140-719-8

Query Match 100.0%; Score 28; DB 10; Length 5;  
Best Local Similarity 100.0%; Pred. No. 4e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYEMK 5  
DB 1 NYEMK 5

RESULT 3  
US-09-731-872-445  
Sequence 445, Application US/09731872  
Patent No. US20020102604A1  
GENERAL INFORMATION:  
APPLICANT: Dumas Milne Edwards, Jean Baptiste  
APPLICANT: Bougueleret, Lydie  
APPLICANT: Jobert, Severin  
TITLE OF INVENTION: FULL-LENGTH HUMAN CDNAS ENCODING POTENTIALLY SECRETED PROTEINS  
FILE REFERENCE: 78 US3 REG  
CURRENT APPLICATION NUMBER: US/09/731,872  
CURRENT FILING DATE: 2000-12-07  
PRIOR APPLICATION NUMBER: US 60/169,629  
PRIOR FILING DATE: 1999-12-08  
PRIOR APPLICATION NUMBER: US 60/187,470  
PRIOR FILING DATE: 2000-03-06  
NUMBER OF SEQ ID NOS: 482  
SOFTWARE: Patent.pm  
SEQ ID NO 445  
LENGTH: 163  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-731-872-445

Query Match 100.0%; Score 28; DB 10; Length 163;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYEMK 5  
DB 159 NYEMK 163

RESULT 4  
US-10-091-442-34  
Sequence 34, Application US/10091442  
Patent No. US20020164711A1  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IMASA, Fuyuki  
APPLICANT: TSUBOKAWA, No. US20020164711A1uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. US20020164711A1uh1ro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, No. US20020164711A1om1  
TITLE OF INVENTION: MEKANARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:



ADDRESSER: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/091,442  
FILING DATE: 07-Mar-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-Aug-1998  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 34:  
US-10-091-442-34  
Query Match 100.0%; Score 28; DB 9; Length 380;  
Best Local Similarity 100.0%; Pred. No. 90;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 NYEMK 5  
Db 284 NYEMK 288

RESULT 5  
US-09-140-719-34  
Sequence 34, Application US/09140719  
Patent No. US20010026931A1  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSUROOKA, NO. US20010026931A1UO  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, NO. US20010026931A1uhlro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAIICHI, Kozo  
APPLICANT: YAMAGUCHI, NO. US20010026931A1oml  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States

ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-Aug-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-140-719-34  
Query Match 100.0%; Score 28; DB 10; Length 380;  
Best Local Similarity 100.0%; Pred. No. 90;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 NYEMK 5  
Db 284 NYEMK 288

RESULT 6  
US-10-029-180-64  
Sequence 64, Application US/10029180  
Publication No. US20020182708A1  
GENERAL INFORMATION:  
APPLICANT: Call, Brian M.  
APPLICANT: Holtzman, Doug  
APPLICANT: Madden, Kevin T.  
APPLICANT: Milna, G. Todd  
APPLICANT: Sherman, Amir  
APPLICANT: Silva, Jeffery C.  
APPLICANT: Trueheart, Josh  
TITLE OF INVENTION: No. US20020182708A1el Regulators of Fungal Gene Expression  
FILE REFERENCE: MIC-004  
CURRENT APPLICATION NUMBER: US/10/029,180  
PRIOR FILING DATE: 2001-12-22  
PRIOR APPLICATION NUMBER: US 60/257,431  
PRIOR FILING DATE: 2000-12-22  
NUMBER OF SEQ ID NOS: 138  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO: 64  
LENGTH: 399  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: fungal gene

US-10-029-180-64

Query Match 100.0%; Score 28; DB 9; Length 399;  
 Best Local Similarity 100.0%; Pred. No. 94;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYEMK 5  
 DB 384 NYEMK 388

RESULT 7

US-09-995-749A-12  
 ; Sequence 12, Application US/09995749A  
 ; Patent No. US20020155568A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
 ; APPLICANT: DIJKHUIZEN, LUBBERT  
 ; APPLICANT: RAHAOUI, HAKIM  
 ; APPLICANT: LEER, ROBERT-JAN  
 ; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES  
 ; FILE REFERENCE: BO43388-CIP  
 ; CURRENT APPLICATION NUMBER: US/09/995,749A  
 ; CURRENT FILING DATE: 2001-11-29  
 ; PRIOR APPLICATION NUMBER: 09/604,957  
 ; PRIOR FILING DATE: 2000-06-28  
 ; PRIOR APPLICATION NUMBER: EPO 00201871.1  
 ; PRIOR FILING DATE: 2000-05-25  
 ; NUMBER OF SEQ ID NOS: 19  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 12  
 ; LENGTH: 584  
 ; TYPE: PRT  
 ; ORGANISM: Leuconostoc mesenteroides  
 US-09-995-749A-12

Query Match 100.0%; Score 28; DB 9; Length 584;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYEMK 5  
 DB 413 NYEMK 417

RESULT 8

US-10-177-293-498  
 ; Sequence 498, Application US/10177293  
 ; Publication No. US20030124128A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lillie, James  
 ; APPLICANT: Glatf, Karen  
 ; APPLICANT: Zhao, Xumel  
 ; APPLICANT: Cannavarpu, Manjula  
 ; APPLICANT: Kamakar, Shubhang  
 ; APPLICANT: Mertens, Maureen  
 ; APPLICANT: Myer, VIC  
 ; APPLICANT: Wang, Youzhen  
 ; APPLICANT: Xu, Yongyao  
 ; APPLICANT: Hoersch, Sebastian  
 ; APPLICANT: Monahan, John  
 ; APPLICANT: Meyers, Rachel E.  
 ; APPLICANT: Bast Jr., Robert C.  
 ; APPLICANT: Hortobagyi, Gabriel N.  
 ; APPLICANT: Puzstai, Lajos  
 ; APPLICANT: Meric, Funda  
 ; APPLICANT: Sahin, Aysegul  
 ; APPLICANT: Mills, Gordon B.  
 ; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,  
 ; FILE REFERENCE: MRI-038  
 ; CURRENT APPLICATION NUMBER: US/10/177,293  
 ; CURRENT FILING DATE: 2002-06-21

PRIOR APPLICATION NUMBER: US 60/299,887  
 ; PRIOR FILING DATE: 2001-06-21  
 ; PRIOR APPLICATION NUMBER: US 60/301,572  
 ; PRIOR FILING DATE: 2001-06-27  
 ; PRIOR APPLICATION NUMBER: US 60/306,501  
 ; PRIOR FILING DATE: 2001-07-18  
 ; PRIOR APPLICATION NUMBER: US 60/325,002  
 ; PRIOR FILING DATE: 2001-09-25  
 ; PRIOR APPLICATION NUMBER: US 60/362,585  
 ; PRIOR FILING DATE: 2002-03-05  
 ; PRIOR APPLICATION NUMBER: US 60/xxx,xxx  
 ; PRIOR FILING DATE: 2002-05-14  
 ; NUMBER OF SEQ ID NOS: 506  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 498  
 ; LENGTH: 847  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-177-293-498

Query Match 100.0%; Score 28; DB 9; Length 847;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYEMK 5  
 DB 438 NYEMK 442

RESULT 9

US-09-861-451A-36  
 ; Sequence 36, Application US/09861451A  
 ; Patent No. US20020068289A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Commonwealth Scientific & Industrial Research Orga  
 ; TITLE OF INVENTION: Methods of Identifying Antigen Gene Sequences  
 ; FILE REFERENCE: FF34033/01  
 ; CURRENT APPLICATION NUMBER: US/09/861,451A  
 ; CURRENT FILING DATE: 2001-05-21  
 ; PRIOR APPLICATION NUMBER: PP7273  
 ; PRIOR FILING DATE: 1998-11-20  
 ; NUMBER OF SEQ ID NOS: 84  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 36  
 ; LENGTH: 299  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: Deduced protein  
 ; OTHER INFORMATION: sequence from clone PAD784  
 US-09-861-451A-36

Query Match 89.3%; Score 25; DB 10; Length 299;  
 Best Local Similarity 80.0%; Pred. No. 3.4e+02;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYEMK 5  
 DB 117 NYEMK 121

RESULT 10

US-09-853-386-69  
 ; Sequence 69, Application US/09853386  
 ; Patent No. US20020049151A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Murphy, Evelyn  
 ; APPLICANT: Bresnahan, Barry  
 ; APPLICANT: Connely, Orla  
 ; APPLICANT: Fitzgerald, Oliver  
 ; TITLE OF INVENTION: Therapeutic Approaches to Diseases by Suppression of the NURR  
 ; FILE REFERENCE: P019720S1

;; CURRENT APPLICATION NUMBER: US/09/853,386  
;; CURRENT FILING DATE: 2001-05-11  
;; PRIOR APPLICATION NUMBER: US 60/203645  
;; PRIOR FILING DATE: 2000-05-12  
;; NUMBER OF SEQ ID NOS: 153  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO 69  
;; LENGTH: 446  
;; TYPE: PRT  
;; ORGANISM: Sus scrofa  
US-09-853-386-69

Query Match 89.3%; Score 25; DB 10; Length 446;  
Best Local Similarity 80.0%; Pred. No. 5e+02;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 NYEMK 5  
Db 68 NYELK 72

RESULT 11  
US-09-878-672-2  
;; Sequence 2, Application US/09878672  
;; Publication No. US20030049812A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Wallis, Nicola G.  
;; TITLE OF INVENTION: NOVEL histidine kinase  
;; NUMBER OF SEQUENCES: 7  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Dechert Price & Rhoads  
;; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
;; CITY: Philadelphia  
;; STATE: PA  
;; COUNTRY: US  
;; ZIP: 19103  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette  
;; OPERATING SYSTEM: IBM Compatible  
;; SOFTWARE: FastSeq for Windows Version 2.0  
;; APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/878,672  
;; FILING DATE: 11-Jun-2001  
;; CLASSIFICATION: <Unknown>  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/916,481  
;; FILING DATE: <Unknown>  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Dickinson, Todd O  
;; REGISTRATION NUMBER: 28,354  
;; REFERENCE/DOCKET NUMBER: GM10022-1  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 215-994-2252  
;; TELEFAX: 215-994-2222  
;; TELEX: <Unknown>  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 554 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-878-672-2

Query Match 89.3%; Score 25; DB 9; Length 554;  
Best Local Similarity 80.0%; Pred. No. 6.2e+02;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 NYEMK 5  
Db 529 NYOMK 533

RESULT 12  
US-09-800-396-4  
;; Sequence 4, Application US/09800396  
;; Patent No. US20020065395A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Wallis, Nicola  
;; TITLE OF INVENTION: NOVEL RESPONSE REGULATOR  
;; NUMBER OF SEQUENCES: 6  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Dechert Price & Rhoads  
;; STREET: 997 Lenox Drive, Building 3, Suite 210  
;; CITY: Lawrenceville  
;; STATE: NJ  
;; COUNTRY: USA  
;; ZIP: 08543

;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette  
;; OPERATING SYSTEM: DOS  
;; SOFTWARE: FastSeq for Windows Version 2.0  
;; APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/800,396  
;; FILING DATE: 06-Mar-2001  
;; CLASSIFICATION: <Unknown>  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/879,531  
;; FILING DATE: <Unknown>  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Bloom, Allen  
;; REGISTRATION NUMBER: 29,135  
;; REFERENCE/DOCKET NUMBER: GM10018  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 609-520-3214  
;; TELEFAX: 609-520-3259  
;; TELEX: <Unknown>  
;; INFORMATION FOR SEQ ID NO: 4:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 554 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
US-09-800-396-4

Query Match 89.3%; Score 25; DB 10; Length 554;  
Best Local Similarity 80.0%; Pred. No. 6.2e+02;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 NYEMK 5  
Db 529 NYOMK 533

RESULT 13  
US-09-878-672-3  
;; Sequence 3, Application US/09878672  
;; Publication No. US20030049812A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Wallis, Nicola G.  
;; TITLE OF INVENTION: NOVEL histidine kinase  
;; NUMBER OF SEQUENCES: 7  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Dechert Price & Rhoads  
;; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
;; CITY: Philadelphia  
;; STATE: PA  
;; COUNTRY: US  
;; ZIP: 19103  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette  
;; OPERATING SYSTEM: DOS

```

; SOFTWARE: FastSeq for Windows Version 2.0
;
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/878,672
;   FILING DATE: 11-Jun-2001
;   CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/916,481
;   FILING DATE: <Unknown>
;
; ATTORNEY/AGENT INFORMATION:
;   NAME: Dickinson, Todd O
;   REGISTRATION NUMBER: 28,354
;   REFERENCE/DOCKET NUMBER: GMI0022-1
;
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 215-994-2252
;   TELEFAX: 215-994-2222
;   TELEX: <Unknown>
;
; INFORMATION FOR SEQ ID NO: 3:
;
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 563 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;
;   SEQUENCE DESCRIPTION: SEQ ID NO: 3:
;
; US-09-878-672-3
;
Query Match      89.3%; Score 25; DB 9; Length 563;
Best Local Similarity 80.0%; Pred. No. 6.3e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY      1 NYEMK 5
Db      538 NTQMK 542

RESULT 14
US-09-853-386-63
; Sequence 63, Application US/09853386
; Patent No. US20020049151A1
; GENERAL INFORMATION:
;   APPLICANT: Murphy, Evelyn
;   APPLICANT: Bresnihan, Barry
;   APPLICANT: Conneely, Orla
;   APPLICANT: Fitzgerald, Oliver
;   TITLE OF INVENTION: Therapeutic Approaches to Diseases by Suppression of the NURR
;   FILE REFERENCE: P019720S1
;   CURRENT APPLICATION NUMBER: US/09/853,386
;   CURRENT FILING DATE: 2001-05-11
;   PRIOR APPLICATION NUMBER: US 60/203645
;   PRIOR FILING DATE: 2000-05-12
;   NUMBER OF SEQ ID NOS: 153
;   SOFTWARE: PatentIn version 3.1
;   SEQ ID NO 63
;     LENGTH: 625
;     TYPE: PRT
;     ORGANISM: HUMAN
;
US-09-853-386-63

Query Match      89.3%; Score 25; DB 10; Length 625;
Best Local Similarity 80.0%; Pred. No. 6.9e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY      1 NYEMK 5
Db      68 NYELK 72

RESULT 15
US-09-853-386-64
; Sequence 64, Application US/09853386
; Patent No. US20020049151A1
; GENERAL INFORMATION:
;   APPLICANT: Murphy, Evelyn
```

```

; APPLICANT: Bresnihan, Barry
; APPLICANT: Conneely, Orla
; APPLICANT: Fitzgerald, Oliver
;   TITLE OF INVENTION: Therapeutic Approaches to Diseases by Suppression of the NURR
;   FILE REFERENCE: P019720S1
;   CURRENT APPLICATION NUMBER: US/09/853,386
;   CURRENT FILING DATE: 2001-05-11
;   PRIOR APPLICATION NUMBER: US 60/203645
;   PRIOR FILING DATE: 2000-05-12
;   NUMBER OF SEQ ID NOS: 153
;   SOFTWARE: PatentIn version 3.1
;   SEQ ID NO 64
;     LENGTH: 626
;     TYPE: PRT
;     ORGANISM: HUMAN
;
US-09-853-386-64

Query Match      89.3%; Score 25; DB 10; Length 626;
Best Local Similarity 80.0%; Pred. No. 6.9e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY      1 NYEMK 5
Db      68 NYELK 72

Search completed: July 11, 2003, 12:37:34
Job time : 1.4657 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: July 11, 2003, 11:45:06 ; Search time 1.00832 Seconds  
(without alignments)  
476.708 Million cell updates/sec

Title: US-10-091-442-8

Perfect score: 28  
Sequence: 1 NYEMK 5

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28	100.0	124	1 MN21B	nonstructural prot
2	28	100.0	124	1 MN21B	nonstructural prot
3	28	100.0	128	1 A49166	3',5'-cyclic-AMP p
4	28	100.0	330	2 D97235	probable hydrolase
5	28	100.0	358	2 T25790	hypothetical prote
6	28	100.0	420	2 B97276	glycosyltransferas
7	28	100.0	479	2 T09365	cytochrome P450 ho
8	28	100.0	538	2 S49933	hypothetical prote
9	28	100.0	608	2 T18445	hypothetical prote
10	28	100.0	900	2 T14277	myosin-like protei
11	28	100.0	2059	2 T41933	large tegument pro
12	25	89.3	62	2 T36921	gene MHC DQ-alpha
13	25	89.3	62	2 T61802	gene MHC DQ-alpha
14	25	89.3	75	2 S40739	hypothetical prote
15	25	89.3	133	2 S12097	rfai protein - Sal
16	25	89.3	182	2 T33023	hypothetical prote
17	25	89.3	193	1 S39401	hydrogenase matura
18	25	89.3	214	2 S45465	heat shock protein
19	25	89.3	230	2 S64784	hypothetical prote
20	25	89.3	235	2 T38333	ABD1 N-acetyl tran
21	25	89.3	265	2 G49508	integrase/recombin
22	25	89.3	267	2 S43900	methyl coenzyme M
23	25	89.3	289	2 E97185	polysaccharide dea
24	25	89.3	300	2 S55620	capsid protein 26
25	25	89.3	300	2 T33232	hypothetical prote
26	25	89.3	335	2 H91191	hypothetical prote
27	25	89.3	335	2 A86039	probable LPS biosy
28	25	89.3	337	2 AH0972	lipopolysaccharide
29	25	89.3	339	2 E83419	probable permease

30	25	89.3	377	2 B64428	formate hydrogently
31	25	89.3	386	2 H90485	glucose 1-dehydrog
32	25	89.3	406	2 S64493	hypothetical prote
33	25	89.3	411	2 C64501	probable phosphono
34	25	89.3	415	2 S12357	interleukin-5 rece
35	25	89.3	416	2 T16058	hypothetical prote
36	25	89.3	416	2 S19896	plasmidogen activa
37	25	89.3	422	2 JC5325	methicillin resist
38	25	89.3	438	2 B83295	hypothetical prote
39	25	89.3	476	2 I37136	adenylate cyclase
40	25	89.3	502	2 I39876	lipoprotein IplA -
41	25	89.3	523	2 F86217	protein T2767.12 l
42	25	89.3	563	2 C97944	histidine kinase (
43	25	89.3	567	2 T34339	hypothetical prote
44	25	89.3	625	2 S71930	neuron-derived rec
45	25	89.3	664	1 S73624	hypothetical prote

#### ALIGNMENTS

RESULT 1  
MNN21B  
nonstructural protein 1B - human respiratory syncytial virus  
C:Species: human respiratory syncytial virus  
C:Date: 30-Jun-1987 #sequence-revision 30-Jun-1987 #text-change 16-Jul-1999  
C:Accession: B94336; C93010; A04032  
R:Collins, P.L.; Wertz, G.W.  
Virology 143, 442-451, 1985  
A:Title: Nucleotide sequences of the 1B and 1C nonstructural protein mRNAs of human r  
A:Reference number: A94336; MUID:86045905; PMID:2998021  
A:Accession: B94336  
A:Molecule type: genomic RNA  
A:Residues: 1-124 <COL>  
A:Cross-references: GB:M11486; GB:K01459; GB:K02719; GB:K03348; GB:K03349; GB:M11217;  
1.1: PID:9333927  
R:Elango, N.; Satake, M.; Venkatesan, S.  
J. Virol. 55, 101-110, 1985  
A:Title: mRNA sequence of three respiratory syncytial virus genes encoding two nonstr  
A:Reference number: A93010; MUID:85237684; PMID:4009789  
A:Molecule type: genomic RNA  
A:Residues: 1-54, 'R', 56-124 <ELA>  
A:Cross-references: GB:M11486  
C:Genetics:  
A:Gene: 1B  
C:Superfamily: respiratory syncytial virus nonstructural protein 1B  
C:Keywords: nonstructural protein

Query Match 100.0%; Score 28; DB 1; Length 124;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NYEMK 5  
Db 64 NYEMK 68

RESULT 2  
MNN21B  
nonstructural protein 1B - human respiratory syncytial virus (strain 16537)  
N:Alternate names: nonstructural protein NS2  
C:Species: human respiratory syncytial virus  
C:Date: 31-Mar-1992 #sequence-revision 31-Mar-1992 #text-change 16-Jun-2000  
C:Accession: B32063  
R:Johnson, P.R.; Collins, P.L.  
J. Gen. Virol. 70, 1539-1547, 1989  
A:Title: The 1B (NS2), 1C (NS1) and N proteins of human respiratory syncytial virus (C  
A:Reference number: A32063; MUID:89279331; PMID:2525176  
A:Accession: B32063  
A:Molecule type: mRNA  
A:Residues: 1-124 <JOH>  
A:Cross-references: EMBL:DD0736; NID:g222559; PIDN:BAA00636.1; PID:g222561

C:Genetics:  
A:Gene: 1B  
C:Superfamily: respiratory syncytial virus nonstructural protein 1B  
C:Keywords: nonstructural protein

Query Match 100.0%; Score 28; DB 1; Length 124;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NTEMK 5  
DB 64 NTEMK 68

## RESULT 3

A49166  
3',5'-cyclic-AMP phosphodiesterase (EC 3.1.4.-), cyclic-GMP inhibited - human (fragments)  
C:Species: Homo sapiens (man)  
C:Date: 03-Feb-1994 #sequence\_revision 03-Feb-1994 #text\_change 19-May-2000  
C:Accession: A49166  
R:Lebon, T.R.; Kasuya, J.; Paxton, R.J.; Belfrage, P.; Hockman, S.; Manganiello, V.C.; F  
Endocrinology 130, 3265-3274, 1992  
A:Title: Purification and characterization of guanosine 3',5'- monophosphate-inhibited 1  
A:Reference number: A49166; MUID:92283180; PMID:1317779  
A:Accession: A49166  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-128 <EB>  
C:Superfamily: cyclic-nucleotide phosphodiesterase, CGMP-inhibited; 3',5'-cyclic-nucleot  
C:Keywords: CGMP binding; phosphoric diester hydrolase

Query Match 100.0%; Score 28; DB 2; Length 128;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NTEMK 5  
DB 17 NTEMK 21

## RESULT 4

D97235  
Probable hydrolase from alpha/beta family, YOKD B. subtilis ortholog [Imported] - Clostr  
C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
C:Accession: D97235  
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
J. Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cl  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: D97235  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-330 <KUR>  
A:Cross-References: GB:AE001437; PIDN:AAK80671.1; PID:g15025760; GSPDB:GN00168  
A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC2725

Query Match 100.0%; Score 28; DB 2; Length 330;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NTEMK 5  
DB 209 NTEMK 213

## RESULT 5

T25790  
hypothetical protein F53E10.5 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T25790  
R:Beck, C.; Wamsley, P.  
submitted to the EMBL Data Library, February 1997  
A:Description: The sequence of C. elegans cosmid F53E10.  
A:Reference number: Z20087  
A:Accession: T25790  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA

A:Residues: 1-358 <BEC>  
A:Cross-References: EMBL:U88177; PIDN:AA842291.1; GSPDB:GN00023; CESP:F53E10.5  
A:Experimental source: strain Bristol N2; clone F53E10  
C:Genetics:  
A:Gene: CESP:F53E10.5  
A:Map position: 5  
A:introns: 23/3; 51/3; 73/1; 103/1; 122/1; 178/1; 191/3; 221/3; 317/3

Query Match 100.0%; Score 28; DB 2; Length 358;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NTEMK 5  
DB 47 NTEMK 51

## RESULT 6

B97276  
glycosyltransferase [Imported] - Clostridium acetobutylicum  
C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
C:Accession: B97276  
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L  
J. Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: B97276  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-420 <KUR>  
A:Cross-References: GB:AE001437; PIDN:AAK80997.1; PID:g15026118; GSPDB:GN00168  
A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC3057

Query Match 100.0%; Score 28; DB 2; Length 420;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NTEMK 5  
DB 191 NTEMK 195

## RESULT 7

T09365  
cytochrome P450 homolog F23K16.120 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 16-Feb-2001  
C:Accession: T09365  
R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.  
submitted to the Protein Sequence Database, June 1999  
A:Reference number: Z16652  
A:Accession: T09365  
A:Molecule type: DNA  
A:Residues: 1-479 <BEV>  
A:Cross-References: EMBL:AL078620; GSPDB:GN00062; ATSP:F23K16.120  
A:Experimental source: cultivar Columbia; BAC clone F23K16  
C:Genetics:  
A:Gene: ATSP:F23K16.120  
A:Map position: 4  
C:Superfamily: Candida cytochrome P450 52A3; cytochrome P450 homology

F:267-446/Domain: cytochrome P450 homology <P45>

Query Match 100.0%; Score 28; DB 2; Length 479;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NYEMK 5  
|||||  
DB 444 NYEMK 448

# RESULT 8

S49933 hypothetical protein YIL045W - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein YI9905.03

C:Species: Saccharomyces cerevisiae

C>Date: 13-Jan-1995 #sequence\_revision 20-Feb-1995 #text\_change 06-Feb-1998

C:Accession: S49933

R:Odeall, C.; Bowman, S.

submitted to the EMBL Data Library, December 1994

A:Reference number: S49931

A:Accession: S49933

A:Molecule type: DNA

A:Residues: 1-538 <ODE>

A:Cross-references: GB:Z47047; EMBL:Z46861; NID:g603997; PID:g763301; MIPS:YIL045W

C:Genetics:

A:Gene: SGD:PIG2

A:Cross-references: SGD:S0001307; MIPS:YIL045W

A:Map position: 9L

C:Superfamily: glucamylase starch-binding domain homology

F:396-516/Domain: glucamylase starch-binding domain homology <SBD>

Query Match 100.0%; Score 28; DB 2; Length 538;  
Best Local Similarity 100.0%; Pred. No. 47;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NYEMK 5  
|||||  
DB 525 NYEMK 529

# RESULT 9

T18445

hypothetical protein C0420W - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 11-Jan-2002

C:Accession: T18445

R:Lawson, D.; Bowman, S.; Barrell, B.

submitted to the EMBL Data Library, August 1997

A:Reference number: Z18935

A:Accession: T18445

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-608 <LAM>

A:Cross-references: EMBL:Z98547; NID:el325376; PIDN:CAB1118.2

C:Genetics:

A:Map position: 3

A:Introns: 326/3; 490/3; 533/3

A:Note: C0420W

Query Match 100.0%; Score 28; DB 2; Length 608;  
Best Local Similarity 100.0%; Pred. No. 53;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NYEMK 5  
|||||  
DB 270 NYEMK 274

# RESULT 10

T14277

myosin-like protein my3 - common sunflower

N:Alternate names: unconventional myosin

C:Species: Helianthus annuus (common sunflower)

C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 23-Mar-2001

C:Accession: T14277

R:Yugrek, O.

submitted to the EMBL Data Library, March 1997

A:Description: Molecular cloning of a new mini myosin from sunflower with similaritie

A:Reference number: Z14226

A:Accession: T14277

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-900 <YUG>

A:Cross-references: EMBL:U94783; NID:g2731701; PID:g2731702

C:Genetics:

A:Gene: MY3

C:Superfamily: myosin motor domain homology

F:168-807/Domain: myosin motor domain homology <MMO>

Query Match 100.0%; Score 28; DB 2; Length 900;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NYEMK 5  
|||||  
DB 280 NYEMK 284

# RESULT 11

T41933

large tegument protein - human herpesvirus 7 (strain J1)

C:Species: human herpesvirus 7

A:Variety: strain J1

C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jan-2000

C:Accession: T41933

R:Nicholas, J.

submitted to the EMBL Data Library, December 1995

A:Description: Determination and analysis of the complete nucleotide sequence of huma

A:Reference number: Z22022

A:Accession: T41933

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2059 <NIC>

A:Cross-references: EMBL:U43400; PIDN:AAC54693.1

C:Genetics:

A:Note: U31

C:Superfamily: varicella-zoster virus 240K tegument protein

Query Match 100.0%; Score 28; DB 2; Length 2059;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NYEMK 5  
|||||  
DB 1182 NYEMK 1186

# RESULT 12

I36921

gene MHC DO-alpha 1 protein - hamadryas baboon (fragment)

C:Species: Papio hamadryas (hamadryas baboon)

C>Date: 07-Jun-1996 #sequence\_revision 07-Jun-1996 #text\_change 23-Jul-1999

C:Accession: I36921

R:Feuter, M.; Otting, N.; Anholts, J.; Leunissen, J.A.M.; Jonker, M.; Bontrop, R.E.

Immunogenetics 36, 71-78, 1992

A:Title: Evolutionary relationships among the primate MHC-DQA1 and DQA2 alleles.

A:Reference number: I36899; MUID:92307745; PMID:1612647

A:Accession: I36921

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-62 <RES>

A:Cross-references: GB:M76226; NID:q176597; PIDN:AAA5391.1; PID:q176598

C:Superfamily: class II histocompatibility antigen; immunoglobulin homology

Query Match 89.3%; Score 25; DB 2; Length 62;  
 Best Local Similarity 80.0%; Pred. No. 28;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 NYEMK 5  
 ||:|  
 Db 56 NYQMK 60

## RESULT 13

I61802  
 gene MHC DQ-alpha 1 protein - rhesus macaque (fragment)  
 C:Species: Macaca mulatta (rhesus macaque)  
 C:Date: 07-Jun-1996 #sequence\_revision 07-Jun-1996 #text\_change 23-Jul-1999  
 C:Accession: I61802  
 R:Kenter, M.; Oetting, N.; Anholts, J.; Leunissen, J.A.M.; Jonker, M.; Bontrop, R.E.  
 Immunogenetics 36, 71-78, 1992  
 A:Title: Evolutionary relationships among the primate MHC-DQA1 and DQA2 alleles.  
 A:Reference number: I36899; MID:92307745; PMID:1612647  
 A:Accession: I61802  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-62 <RSS>  
 A:Cross-references: GB:M76229; NID:9342161; PIDN:AAA36870.1; PID:9342162  
 C:Superfamily: class II histocompatibility antigen; immunoglobulin homology

## Query Match

89.3%; Score 25; DB 2; Length 62;  
 Best Local Similarity 80.0%; Pred. No. 28;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYEMK 5  
 ||:|

Db 56 NYQMK 60

## RESULT 14

S40739  
 hypothetical protein ZK507.5 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 09-Sep-1997  
 C:Accession: S40739  
 R:Hawkins, T.; Thomas, K.  
 submitted to the EMBL Data Library, December 1993  
 A:Reference number: S40735  
 A:Accession: S40739  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-75 <HAM>  
 A:Cross-references: EMBL:229116; NID:91067087; PID:9439269  
 C:Genetics:  
 A:introns: 32/1

## Query Match

89.3%; Score 25; DB 2; Length 75;  
 Best Local Similarity 80.0%; Pred. No. 35;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYEMK 5  
 ||:|

Db 47 NYEMR 51

## RESULT 15

S12097  
 rfaI protein - Salmonella typhimurium  
 C:Species: Salmonella typhimurium  
 C:Date: 21-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 08-Oct-1999  
 C:Accession: S12097  
 R:Carstenius, P.; Flock, J.I.; Lindberg, A.  
 Nucleic Acids Res. 18, 6128, 1990  
 A:Title: Nucleotide sequence of rfaI and rfaJ genes encoding lipopolysaccharide glycosyl  
 A:Reference number: S12097; MID:91045080; PMID:2235496  
 A:Accession: S12097  
 A:Status: preliminary

A:Molecule type: DNA  
 A:Residues: 1-133 <CAR>  
 A:Cross-references: EMBL:X53847; NID:947882; PIDN:CAA37841.1; PID:947883

## Query Match

89.3%; Score 25; DB 2; Length 133;  
 Best Local Similarity 80.0%; Pred. No. 62;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYEMK 5  
 ||:|

Db 41 NYELK 45

Search completed: July 11, 2003, 12:00:42  
 Job time : 2.10832 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:43:41 ; Search time 0.47817 Seconds

(without alignments)  
433.698 Million cell updates/sec

Title: US-10-091-442-8

Perfect score: 28

Sequence: 1 NYEMK 5

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt\_40.\*

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	28	100.0	124	VNS2_BRSVA	Q65695 bovine resp
2	28	100.0	124	VNS2_HRSVA	P24569 human resp
3	28	100.0	124	VNS2_HRSVA	P04543 human resp
4	28	100.0	124	VNS2_HRSVA	Q42038 human resp
5	28	100.0	124	VNS2_HRSVA	Q86305 human resp
6	28	100.0	124	VNS2_HRSVA	Q65707 ovine resp
7	28	100.0	380	SPB7_HUMAN	Q75635 homo sapien
8	28	100.0	538	PIG2_YEAST	P40187 saccharomyc
9	28	100.0	847	VAV3_HUMAN	Q9UK44 homo sapien
10	28	100.0	847	VAV3_HUMAN	Q9UK44 homo sapien
11	28	100.0	2059	TEGU_HSV7J	P52362 human herpe
12	28	100.0	149	UREE_LACFE	Q9X525 lactobacill
13	28	100.0	193	HUPD_BRAJA	Q45251 bradyrhizob
14	28	100.0	203	ARHD_DICDI	P36416 dictyostel
15	28	100.0	213	HS26_YEAST	P15992 saccharomyc
16	28	100.0	235	ARHD_HUMAN	P41227 homo sapien
17	28	100.0	337	FEAT_SALTY	P19816 salmonella
18	28	100.0	406	YGA3_YEAST	P53298 saccharomyc
19	28	100.0	411	APG1_METJA	Q59007 methanococc
20	28	100.0	415	ILSR_MOUSE	P21183 mus musculu
21	28	100.0	416	PAIR_RAT	P29524 rattus norv
22	28	100.0	502	LPLA_BACSU	P37966 bacillus su
23	28	100.0	626	NRA3_HUMAN	Q92570 homo sapien
24	28	100.0	664	Y366_MYCPN	F75234 mycoplasma
25	28	100.0	666	SPOT_AOUAE	O67012 aquilex aeo
26	28	100.0	762	SLAP_ACEKI	P22258 acetogenium
27	28	100.0	880	YL86_YEAST	Q06708 saccharomyc
28	28	100.0	887	CYAA_HUMAN	Q08462 homo sapien
29	28	100.0	906	SECA_RICPR	O9ZC97 rickettsia
30	28	100.0	1090	CYAA_RAT	P26769 rattus norv
31	28	100.0	1131	MYF_CHICK	P16419 gallus gall
32	28	100.0	1188	S3B1_SCHPO	Q10178 schizosacch
33	28	100.0	1386	RPOD_MARPO	P06274 marchantia

34	24	85.7	113	1	RL22_BACHD	Q929K9 bacillus ha
35	24	85.7	157	1	GRI_ECOLI	P31012 escherichia
36	24	85.7	178	1	AR21_HUMAN	P15145 homo sapien
37	24	85.7	182	1	GRF_MOUSE	O88593 mus musculu
38	24	85.7	246	1	Y181_METJA	O57640 methanococc
39	24	85.7	252	1	UDP_ECOLI	P12758 escherichia
40	24	85.7	252	1	UDP_ECOLI	O33808 salmonella
41	24	85.7	253	1	UDP_KLEAE	O08444 klebsiella
42	24	85.7	256	1	Y165_METJA	O57629 methanococc
43	24	85.7	259	1	UDP_HAETN	P43770 haemophilus
44	24	85.7	268	1	PLSC_MYCCE	Q49402 mycoplasma
45	24	85.7	274	1	AROK_PYRAB	Q9V1H6 pyrococcus

## ALIGNMENTS

RESULT 1	VNS2_BRSVA	STANDARD;	PRT;	124 AA.
AC	Q65695;			
DT	15-JUL-1998 (Rel. 36, Created)			
DT	15-JUL-1998 (Rel. 36, Last sequence update)			
DT	15-JUL-1998 (Rel. 36, Last annotation update)			
DE	Nonstructural protein 2 (Nonstructural protein 1B).			
GN	1B OR NS2.			
OS	Bovine respiratory syncytial virus (strain A51908) (BRS).			
OC	Viruses; ssRNA negative-strand viruses; Mononegavirales;			
OC	Paramyxoviridae; Pneumovirinae; Pneumovirus.			
OX	NCBI_TaxID=11247;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=95146950; PubMed=7844532;			
RA	Pasteur M.K., Samal S.K.;			
RT	"Nucleotide sequence analysis of the non-structural NS1 (1C) and NS2			
RT	(1B) protein genes of bovine respiratory syncytial virus.";			
RL	J. Gen. Virol. 76:193-197(1995).			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; U15938; AAA85672.1; -			
DR	InterPro; IPR004336; RSV_NS2.			
DR	Pfam; PF03113; RSV_NS2; 1.			
KW	Nonstructural protein.			
SO	SEQUENCE 124 AA; 14576 MW; B47548BD5FB8FFD CRC64;			
Query Match	100.0%; Score 28; DB 1; Length 124;			
Best Local Similarity	100.0%; Pred. No. 6.1;			
Matches	5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
OY	1 NYEMK 5			
DB	64 NYEMK 68			
RESULT 2	VNS2_HRSVA	STANDARD;	PRT;	124 AA.
ID	VNS2_HRSVA			
AC	P24569;			
DT	01-MAR-1992 (Rel. 21, Created)			
DT	01-MAR-1992 (Rel. 21, Last sequence update)			
DT	01-MAR-1992 (Rel. 21, Last annotation update)			
DE	Nonstructural protein 2 (Nonstructural protein 1B).			
GN	1B OR NS2.			
OS	Human respiratory syncytial virus (subgroup B / strain 18537).			
OC	Viruses; ssRNA negative-strand viruses; Mononegavirales;			
OC	Paramyxoviridae; Pneumovirinae; Pneumovirus.			

```

OX NCB1_TaxID-11251;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-89279331; PubMed-2525176;
RA Johnson P.R., Collins P.L.;
RT "The 1B (NS2), 1C (NS1) and N proteins of human respiratory syncytial
RT virus (RSV) of antigenic subgroups A and B: sequence conservation and
RT divergence within RSV genomic RNA.";
RL J. Gen. Virol. 70:1539-1547(1989).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D00736; BAA00636.1; -
DR PIR: B32063; MNZ215;
DR InterPro: IPR004336; RSV_NS2.
DR Pfam: PF03113; RSV_NS2; 1.
DR Nonstructural protein.
KW SEQUENCE 124 AA; 14624 MW; 526F8C2C6553C1DA CRC64;

Query Match 100.0%; Score 28; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NYEMK 5
DB 64 NYEMK 68

RESULT 3
VNS2_HRSVA
ID VNS2_HRSVA STANDARD; PRT; 124 AA.
AC P04543;
DT 13-AUG-1987 (Rel. 05, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Nonstructural protein 2 (Nonstructural protein 1B).
GN 1B OR NS2.
OS Human respiratory syncytial virus (strain A2).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Pneumovirinae; Pneumovirus.
OX NCB1_TaxID-11259;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-86045905; PubMed-2998021;
RA Collins P.L., Wertz G.W.;
RT "Nucleotide sequences of the 1B and 1C nonstructural protein mRNAs of
RT human respiratory syncytial virus.";
RL Virology 143:442-451(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-85237684; PubMed-4009789;
RA Elango N., Satake M., Venkatesan S.;
RT "mRNA sequence of three respiratory syncytial virus genes encoding
RT two nonstructural proteins and a 22k structural protein.";
RL J. Virol. 55:101-110(1985).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE-95266253; PubMed-7747420;
RA Connors M., Crowe J.E. Jr., Firestone C.Y., Murphy B.R., Collins P.L.;
RT "A cold-passaged, attenuated strain of human respiratory syncytial
RT virus contains mutations in the F and L genes.";
RL Virology 208:478-484(1995).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE-97187925; PubMed-9035372;
RA Crowe J.E. Jr., Firestone C.Y., Whitehead S.S., Collins P.L.,
RA Murphy B.R.;

```

```

RT "Acquisition of the ts phenotype by a chemically mutagenized cold-
RT passed human respiratory syncytial virus vaccine candidate results
RT from the acquisition of a single mutation in the polymerase (L)
RT gene.";
RL Virus Genes 13:269-273(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M11486; AAB59851.1; -
DR EMBL: U50362; AAB86657.1; -
DR EMBL: U50363; AAB86669.1; -
DR EMBL: U63644; AAC55963.1; -
DR PIR: A04032; MNZ21B;
DR InterPro: IPR004336; RSV_NS2.
DR Pfam: PF03113; RSV_NS2; 1.
DR Nonstructural protein.
FT CONFLICT 55 R -> K (IN REF. 1).
KW SEQUENCE 124 AA; 14702 MW; C14DF0E9C9CE5512 CRC64;

Query Match 100.0%; Score 28; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NYEMK 5
DB 64 NYEMK 68

RESULT 4
VNS2_HRSVB
ID VNS2_HRSVB STANDARD; PRT; 124 AA.
AC Q42038;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Nonstructural protein 2 (Nonstructural protein 1B).
GN 1B OR NS2.
OS Human respiratory syncytial virus (strain B1).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Pneumovirinae; Pneumovirus.
OX NCB1_TaxID-79692;
RN [1]
RP SEQUENCE FROM N.A.
RA Karon R.A., Buonagurio D.A., Georgiu A.F., Whitehead S.S.,
RA Adamus J.E., Clements-Mann M.L., Harris D.O., Randolph V.B.,
RA Udem S.A., Murphy B.R., Sidhu M.S.;
RT Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RL
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF013254; AAB82430.1; -
DR EMBL: AF013255; AAB82441.1; -
DR InterPro: IPR004336; RSV_NS2.
DR Pfam: PF03113; RSV_NS2; 1.
DR Nonstructural protein.
KW SEQUENCE 124 AA; 14567 MW; A9FD37D964F97073 CRC64;

Query Match 100.0%; Score 28; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy      1 NFEWK 5
       11111
Db      64 NFEWK 68

RESULT 5
VNS2_HRSVL STANDARD: PRT: 124 AA.
ID VNS2_HRSVL
AC 086305:
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 15-JUL-1998 (Rel. 36, Last annotation update)
DE Nonstructural protein 2 (Nonstructural protein 1B).
GN 1B OR NS2.
OS Human respiratory syncytial virus (subgroup A / strain Long).
OC Viruses: ssRNA negative-strand viruses; Mononegavirales.
OC Paramyxoviridae; Pneumovirinae; Pneumovirus.
OX NCBI_TaxID=11260;
RN [1]
RP SEQUENCE FROM N.A.
RA Submitted (Aug-1995) to the EMBL/GenBank/DBJ databases.
RL -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: U35029; AAA79090.1; -
DR InterPro: IPR004336; RSV_NS2.
DR Pfam: PF03113; RSV_NS2; 1.
KM Nonstructural protein.
SQ SEQUENCE 124 AA; 14705 MW; 98657318FEB744E1 CRC64;

Query Match 100.0%; Score 28; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 NFEWK 5
       11111
Db      64 NFEWK 68

RESULT 6
VNS2_ORSVN STANDARD: PRT: 124 AA.
ID VNS2_ORSVN
AC 065707:
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Nonstructural protein 2 (Nonstructural protein 1B).
GN 1B.
OS Ovine respiratory syncytial virus (strain MSU 83-1578) (ORSV).
OC Viruses: ssRNA negative-strand viruses; Mononegavirales.
OC Paramyxoviridae; Pneumovirinae; Pneumovirus.
OX NCBI_TaxID=19699;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-94157498; PubMed-8113762;
RA Alenxari H.M., Potgieter L.N.D.;
RT "Nucleotide and predicted amino acid sequence analysis of the ovine
RT respiratory syncytial virus non-structural 1C and 1B genes and the
RT small hydrophobic protein gene.";
RL J. Gen. Virol. 75:401-404(1994).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: L15451; AAA42813.1; -
DR InterPro: IPR004336; RSV_NS2.
DR Pfam: PF03113; RSV_NS2; 1.
KM Nonstructural protein.
SQ SEQUENCE 124 AA; 14496 MW; 201D0F5DFEA9A4EB CRC64;

Query Match 100.0%; Score 28; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 NFEWK 5
       11111
Db      64 NFEWK 68

RESULT 7
SPB7_HUMAN STANDARD: PRT: 380 AA.
ID SPB7_HUMAN
AC 075635:
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Megsin (TRP55) (Serpin B7).
GN SERPINB7.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97326116; PubMed=9182567;
RA Tsujimoto M., Tsuruoka N., Ishida N., Kurihara T., Iwasa F.,
RA Yamashiro K., Rogi T., Kodama S., Katsuragi N., Adachi M.,
RA Katayama T., Nakao M., Yamachi K., Hashino J., Haruyama M., Miura K.,
RA Nakanishi T., Nakazato H., Teramura M., Mizoguchi H., Yamaguchi N.;
RT "Purification, cDNA cloning, and characterization of a new serpin with
RT megakaryocyte maturation activity.";
RL J. Biol. Chem. 272:15373-15380(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98376492; PubMed=9710452;
RA Miyata T., Nangaku M., Suzuki D., Inagi R., Uragami K., Sakai H.,
RA Okubo K., Kurokawa K.;
RT "A mesangium-predominant gene, megin, is a new serpin upregulated in
RT Iga nephropathy.";
RL J. Clin. Invest. 102:828-836(1998).
CC -!- FUNCTION: Might function as an inhibitor of Lys-specific
CC proteases. Might influence the maturation of megakaryocytes via
CC its action as a serpin.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- TISSUE SPECIFICITY: Predominantly expressed in mesangial cells.
CC -!- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: D88575; BAA31232.1; -
DR EMBL: AF027866; AAC64506.1; -
DR HSSP: P05619; 1HLE.
DR GeneW: HGNC:13902; SERPINB7.
DR MIM: 603357; -
DR InterPro: IPR000215; Serpin.
DR Pfam: PF00079; serpin; 1.
DR SMART: SM00093; SERPIN; 1.

```

DR PROSITE: PS00284; SERPIN; 1.  
 KW Serpin; Serine protease inhibitor.  
 FT ACT\_SITE 347 REACTIVE\_BOND (BY SIMILARITY).  
 SQ SEQUENCE 380 AA; 42904 MW; 9A2CDB6C63CFF605 CRC64;  
 Query Match 100.0%; Score 28; DB 1; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 NTEMK 5  
 DB 284 NTEMK 288

RESULT 8  
 ID PIG2\_YEAST STANDARD; PRT; 538 AA.  
 AC P40187;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE GSY2 interacting protein PIG2.  
 GN PIG2 OR YII045W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=528BC / AB972;  
 RA Barrett B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,  
 RA Church C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,  
 RA Genies S., Hamlyn N., Horsnell T.S., Hunt S., Jagals K., Jones M.,  
 RA Louis E., Lye G., Moule S., Odeli C., Pearson D.,  
 RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,  
 RA Walsh S.V., Whitehead S.;  
 RL Submitted (Dec-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP PARTIAL SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RA MEDLINE=97197965; PubMed=9046081;  
 RA Cheng C., Huang D., Roach P.J.;  
 RT "Yeast PIG genes: PIG1 encodes a putative type 1 phosphatase subunit  
 RT that interacts with the yeast glycogen synthase Gsy2p."  
 RL Yeast 13:1-8(1997).  
 CC -1- FUNCTION: INTERACTS WITH GLYCOGEN SYNTHASE 2 (GSY2); POSSIBLY ALSO  
 CC INTERACTS WITH PHOSPHATASE 1 (GLC7).  
 CC -1- SIMILARITY: TO YEAST GIP2.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage -by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: Z46861; CA86906.1; -;  
 DR SGD: S0001307; PIG2.  
 DR InterPro: IPR005036; CBM\_21.  
 DR Pfam: PF03370; CBM\_21; 1.  
 SQ SEQUENCE 538 AA; 61938 MW; 73ED8F5CC6692172 CRC64;

Query Match 100.0%; Score 28; DB 1; Length 538;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 NTEMK 5  
 DB 525 NTEMK 529

RESULT 9  
 VAV3\_HUMAN

ID VAV3\_HUMAN STANDARD; PRT; 847 AA.  
 AC O9UKW4; O95230; O9Y5X8;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Vav-3 protein.  
 GN VAV3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
 RX MEDLINE=99455043; PubMed=10523675;  
 RA Movilla N., Bustelo X.R.;  
 RT "Biological and regulatory properties of Vav-3, a new member of the  
 RT Vav family of oncoproteins."  
 RL Mol. Cell. Biol. 19:7870-7885(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Breast, and Colon carcinoma;  
 RX MEDLINE=98371222; PubMed=9705494;  
 RA Tremble T., Welsh J., Jung B., Mathieu-Daude F., McClelland M.;  
 RT "Non-stoichiometric reduced complexity probes for cDNA arrays."  
 RL Nucleic Acids Res. 26:3883-3891(1998).  
 CC -1- FUNCTION: EXCHANGE FACTOR FOR GTP-BINDING PROTEINS RHOA, RHOG AND,  
 CC TO A LESSER EXTENT, RAC-1. BINDS PHYSICALLY TO THE NUCLEOTIDE-FREE  
 CC STATES OF THOSE GTPASES.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS, ALPHA (SHOWN HERE) AND BETA;  
 CC ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- SIMILARITY: CONTAINS 1 CALPONIN-HOMOLOGY (CH) DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 DEL-HOMOLOGY (DH) DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 ZINC-DEPENDENT PHORBOL-ESTER AND DAG  
 CC BINDING DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 2 SH3 DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage -by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF118887; AAD20349.1; -;  
 DR EMBL: AF118886; AAD20348.1; -;  
 DR EMBL: AF067817; AAC79695.1; -;  
 DR HSSP: P29355; ISEM.  
 DR Genew: HGNC:12659; VAV3.  
 DR MIM: 605541; -;  
 DR InterPro: IPR003247; CH\_type.  
 DR InterPro: IPR001715; Calponin-like.  
 DR InterPro: IPR002219; DAG\_PE-bind.  
 DR InterPro: IPR001311; GDS\_CDC24.  
 DR InterPro: IPR001849; PH.  
 DR InterPro: IPR000219; RhogEF.  
 DR InterPro: IPR000980; SH2.  
 DR InterPro: IPR001452; SH3.  
 DR InterPro: IPR003096; SM22\_calponin.  
 DR Pfam: PF00017; SH2; 1.  
 DR Pfam: PF00018; SH3; 1.  
 DR Pfam: PF00130; DAG\_PE-bind; 1.  
 DR Pfam: PF00169; PH; 1.  
 DR Pfam: PF00307; CH; 1.  
 DR Pfam: PF00621; RhogEF; 1.  
 DR PRINTS: PR00401; SH2DOMAIN.  
 DR PRINTS: PR00452; SH3DOMAIN.  
 DR PRINTS: PR00888; SM22CALPONIN.  
 DR PRODOM: PD000066; SH3; 1.  
 DR PRODOM: PD000093; SH2; 1.  
 DR PRODOM: PD001527; CH\_type; 1.

```

RESULT 10
VAV3_MOUSE
ID VAV3_MOUSE STANDARD: PRT: 847 AA.
AC Q9R0C6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Vav-3 protein.
GN VAV3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20179693; PubMed=10713454;
RA Trenkle T., McClelland M., Adlrofer K., Welsh J.;
RT "Major transcript variants of VAV3, a new member of the VAV family of
RL guanine nucleotide exchange factors."
RL Gene 245:139-149(2000).
CC -1- FUNCTION: EXCHANGE FACTOR FOR GTP-BINDING PROTEINS RHOA, RHOG AND,
CC TO A LESSER EXTENT, RAC-1. BINDS PHYSICALLY TO THE NUCLEOTIDE-FREE
CC STATES OF THOSE GTPASES (BY SIMILARITY).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; ALPHA (SHOWN HERE) AND BETA;
CC ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: CONTAINS 1 CALPONIN-HOMOLOGY (CH) DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY (DH) DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 ZINC-DEPENDENT PHOSBOYL-ESTER AND DAG
CC BINDING DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.

```

ID	TEGU_HSV/J	STANDARD;	PRI;
AC	P52362;		
DT	01-OCT-1996	(Rel. 34, Created)	

```

Query Match          100.0%; Score 28; DB 1; Length 847;
Best Local Similarity 100.0%; Pred. No. 44;
Matches      5; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

OY      1 NYEMK 5
         |||||
Db      438 NYEMK 442

RESULT 11
TEGU_HSV7J
ID      TEGU_HSV7J      STANDARD:      PRT: 2059 AA.
AC      P52362;
DT      01-OCT-1996 (Rel. 34, Created)

```

```

DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Large tegument protein.
GN U31.
OC Human herpesvirus (type 7 / strain J1) (HHV7).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=57278;
RN [1]
RP SEQUENCE FROM N.A.
RA Nicholas J.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: TEGUMENT PROTEIN.
CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,
CC EHV-1 24, EBV BFLF1, HSV-1 64, VZV 22, AND HCMV UL48.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U43400; AAC54693.1; -
SQ SEQUENCE 2059 AA; 239476 MW; AA6C6ADEDC5D316 CRC64;

Query Match 100.0%; Score 28; DB 1; Length 2059;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 NYEMK 5
Db 1182 NYEMK 1186

RESULT 12
IDREE_LACFE
IDREE_LACFE STANDARD; PRT; 149 AA.
AC Q9X525;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Urease accessory protein ureE.
DE UREASE.
GN Lactobacillus fermentum.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1613;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-TK1214;
RC Cocon E.; Visser J.J.; Van Vuuren H.J.J.;
RL "Urease operon of Lactobacillus fermentum.";
RT Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN UREASE METALLOCENTER ASSEMBLY. BINDS NICKEL.
CC PROBABLY FUNCTIONS AS A NICKEL DONOR DURING METALLOCENTER ASSEMBLY
CC (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE UREE FAMILY.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AF120718; AAD22481.1; -
DR InterPro; IPR004029; UreeE.
DR Pfam; PF02814; Uree; 1.
DR Nickel.

```

```

SO SEQUENCE 149 AA; 17041 MW; 419F74BC0BF414 CRC64;
Query Match 89.3%; Score 25; DB 1; Length 149;
Best Local Similarity 80.0%; Pred. No. 39;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 NYEMK 5
DB 130 NYELK 134

RESULT 13
HUPD_BRAJA STANDARD: PRT: 193 AA.
ID HUPD_BRAJA Q45251; Q45249;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hydrogenase expression/formaton protein hupd.
DE HUPD.
CN Bradyrhizobium japonicum.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Bradyrhizobium group; Bradyrhizobium.
NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94047099; PubMed=8230323;
RA van Soom C., Browaeys J., Verreth C., Vanderleyden J.;
RT "Nucleotide sequence analysis of four genes, hupC, hupF and
RT hupG, downstream of the hydrogenase structural genes in
RT Bradyrhizobium japonicum."
RL J. Mol. Biol. 234:508-512(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=USDA 110;
RX MEDLINE=94215887; PubMed=8163174;
RA Fu C., Water R.J.;
RT "Sequence and characterization of three genes within the hydrogenase
RT gene cluster of Bradyrhizobium japonicum."
RL Gene 141:47-52(1994).
CC -1- FUNCTION: NOT KNOWN. COULD BE INVOLVED IN THE PROCESSING OF
CC HYDROGENASE.
CC
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M52.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/).
CC -----
CC EMBL: Z21948; CAA79944.1; -
CC EMBL: L24446; AAD13469.1; -
CC HSP: P37182; ICFZ.
CC MEROPS: M52.001; -.
CC InterPro: IPR004419; HUPD.
CC InterPro: IPR000671; Hydrgn_uplake.
CC Pfam: PF01750; HycI. 1.
CC PRINTS: PR00446; HYDRGNUPAKE.
CC TIGRfams: TIGR00072; hydrgn_prot. 1.
CC TIGRfams: TIGR00140; hupd. 1.
CC HydroLase; Metalloprotease; Nickel.
CC METAL 23
CC METAL 23
CC METAL 69
CC METAL 100
CC METAL 100
CC CONFLICT 149
CC CONFLICT 149
CC CONFLICT 171
CC SEQUENCE 193 AA; 21429 MW; 34A8A2B53176ADAD CRC64;

Query Match 89.3%; Score 25; DB 1; Length 193;
Best Local Similarity 80.0%; Pred. No. 50;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 NYEMK 5  
 DB 187 NYEMK 191

RESULT 14  
 ARDH.DICDI STANDARD; PRT: 203 AA.

AC P36416: 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE N-terminal acetyltransferase complex ARD1 subunit homolog.  
 GN NTA OR ARD1.  
 OS Dictyostelium discoideum (Slime mold).  
 OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostellum.  
 OX NCBI\_TaxID=44689;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AX2;  
 RA Mueller-Taubenberger A.;  
 RL Submitted (FEB-1994) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: SEEMS TO BE INVOLVED IN N-ACETYLTATION (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE ACETYLTTRANSFERASE FAMILY. ARD1  
 CC SUBFAMILY.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC  
 CC DR EMBL; U06453; AAA16510.1; -.  
 CC DR DictyDb; DD05073; nata.  
 CC DR InterPro; IPR000182; GCN5acetyltransf.  
 CC DR Pfam; PF00583; Acetyltransf; 1.  
 CC KM Transferrase; Acyltransferase.  
 CC SQ SEQUENCE 203 AA; 23415 MW; 501B6B46DF0F146 CRC64;

Query Match 89.3%; Score 25; DB 1; Length 203;  
 Best Local Similarity 80.0%; Pred. No. 53;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 NYEMK 5  
 DB 26 NYOMK 30

RESULT 15  
 HS26\_YEAST STANDARD; PRT: 213 AA.

AC P15992; 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 16, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Heat shock protein 26 (26 kDa heat shock protein).  
 GN HSP26 OR YBR072W OR YBR0714.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=30097950; Pubmed=2689876;  
 RA Susek R.E.; Lindquist S.L.;  
 RT "hsp26 of Saccharomyces cerevisiae is related to the superfamily of  
 RT small heat shock proteins but is without a demonstrable function."  
 RL Mol. Cell. Biol. 9:5265-5271(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.

RX MEDLINE=89378758; Pubmed=2673926;  
 RA Bossier P.; Fitch I.T.; Boucherie H.; Tuile M.F.;  
 RT "Structure and expression of a yeast gene encoding the small  
 RT heat-shock protein Hsp26.";  
 RL Gene 78:323-330(1989).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288c;  
 RX MEDLINE=95076715; Pubmed=7985423;  
 RA van der Aart O.J.M.; Barthe C.; Daignon F.; Aigle M.; Crouzet M.;  
 RA Steensma H.Y.;  
 RT "Sequence analysis of a 31 kb DNA fragment from the right arm of  
 RT Saccharomyces cerevisiae chromosome II.";  
 RL Yeast 10:959-964(1994).  
 CC -1- FUNCTION: NOT KNOWN. ONE OF THE MAJOR POLYPEPTIDES PRODUCED ON  
 CC HEAT SHOCK.  
 CC -1- SUBUNIT: PRESENT IN LARGE COMPLEXES.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING THE ENTRY INTO STATIONARY  
 CC PHASE RESULTING FROM GLUCOSE LIMITATION.  
 CC -1- INDUCTION: BY HEAT SHOCK, AND UNDER OTHER CONDITIONS OF STRESS,  
 CC SUCH AS INCREASED SALT CONCENTRATION AND STARVATION.  
 CC -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)  
 CC FAMILY.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC  
 CC DR EMBL; M23871; AAA66914.1; -.  
 CC DR EMBL; M26942; AAA79010.1; -.  
 CC DR EMBL; X76294; CAA53929.1; -.  
 CC DR EMBL; Z35941; CAA85016.1; -.  
 CC DR PIR; S45465; S45465.  
 CC DR PIR; S39222; S39222.  
 CC DR SGD; S0000276; HSP26.  
 CC DR InterPro; IPR002068; Hsp20.  
 CC DR Pfam; PF00011; HSP20; 1.  
 CC DR PROSITE; PS01031; HSP20; 1.  
 CC KM Heat shock.  
 CC FT INIT.MET 0  
 CC FT CONFLICT 31 31 G -> A (IN REF. 2).  
 CC FT CONFLICT 206 206 S -> C (IN REF. 2).  
 CC SQ SEQUENCE 213 AA; 23748 MW; 1C9C4B0D0626B6A4 CRC64;

Query Match 89.3%; Score 25; DB 1; Length 213;  
 Best Local Similarity 80.0%; Pred. No. 56;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 NYEMK 5  
 DB 102 NYELK 106

Search completed: July 11, 2003, 11:55:20  
 Job time : 2.47817 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: July 11, 2003, 11:44:31; Search time 2.18295 Seconds  
(without alignments)  
471.946 Million cell updates/sec

Title: US-10-091-442-8

Perfect score: 28

Sequence: 1 NYEMK 5

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: 1: SP-archaea:\*  
2: SP-bacteria:\*  
3: SP-fungi:\*  
4: SP-human:\*  
5: SP-invertebrate:\*  
6: SP-mammal:\*  
7: SP-mhc:\*  
8: SP-organellar:\*  
9: SP-phage:\*  
10: SP-plant:\*  
11: SP-rodent:\*  
12: SP-virus:\*  
13: SP-vertebrate:\*  
14: SP-unclassified:\*  
15: SP-virus:\*  
16: SP-bacteriophage:\*  
17: SP-archaeal:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	28	100.0	10 4 09UCP3	09UCP3 homo sapien
2	28	100.0	124 12 091859	091859 bovine resp
3	28	100.0	124 12 09YWM9	09YWM9 bovine resp
4	28	100.0	124 12 09YWM8	09YWM8 bovine resp
5	28	100.0	124 12 09YWM6	09YWM6 bovine resp
6	28	100.0	124 12 09YWM4	09YWM4 bovine resp
7	28	100.0	124 12 09YWM2	09YWM2 bovine resp
8	28	100.0	124 12 09YWM1	09YWM1 bovine resp
9	28	100.0	133 2 09AN06	09AN06 bradyrhizob
10	28	100.0	320 11 089012	089012 mus musculu
11	28	100.0	325 11 089010	089010 mus musculu
12	28	100.0	330 16 097FL1	097FL1 clostridium
13	28	100.0	358 5 091317	091317 caenorhabd1
14	28	100.0	420 2 092FG7	092FG7 staphylococ
15	28	100.0	420 2 0956V4	0956V4 staphylococ
16	28	100.0	420 16 097EQ1	097EQ1 clostridium

17	28	100.0	479 10 09SV49	09SV49 arabidopsis
18	28	100.0	505 5 09NAN7	09NAN7 caenorhabd1
19	28	100.0	562 5 09NUN9	09NUN9 plasmodium
20	28	100.0	580 5 077333	077333 plasmodium
21	28	100.0	702 16 08RBF9	08RBF9 thermomater
22	28	100.0	827 13 08UME6	08UME6 tetradon n
23	28	100.0	846 13 08UX6	08UX6 gallus gall
24	28	100.0	900 10 050004	050004 helianthus
25	28	100.0	1212 5 09UOL0	09UOL0 plasmodium
26	28	100.0	1442 11 08RA22	08RA22 mus musculu
27	28	100.0	2057 2 09RE05	09RE05 leucostoc
28	28	100.0	2059 12 056278	056278 human herpe
29	28	100.0	62 7 030881	030881 papio hamad
30	28	100.0	62 7 030691	030691 macaca mule
31	25	89.3	121 16 08XME1	08XME1 clostridium
32	25	89.3	127 2 070851	070851 borrelia bu
33	25	89.3	133 2 09S0G1	09S0G1 borrelia bu
34	25	89.3	133 2 09R384	09R384 borrelia bu
35	25	89.3	133 2 09R2M2	09R2M2 borrelia bu
36	25	89.3	168 16 08UKA2	08UKA2 agrobacteri
37	25	89.3	182 5 061219	061219 caenorhabd1
38	25	89.3	192 10 09FK14	09FK14 arabidopsis
39	25	89.3	196 5 09VT75	09VT75 drosophila
40	25	89.3	197 10 094F76	094F76 zea mays (m
41	25	89.3	220 11 09C0X6	09C0X6 mus musculu
42	25	89.3	225 11 08R2U5	08R2U5 mus musculu
43	25	89.3	229 4 09BSU3	09BSU3 homo sapien
44	25	89.3	230 3 007843	007843 saccharomyc
45	25	89.3	235 11 09QY36	09QY36 mus musculu

## ALIGNMENTS

RESULT 1  
ID 09UCP3 PRELIMINARY; PRT; 10 AA.  
AC 09UCP3;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE CGMP-inhibited LOW K(M) CAMP phosphodiesterase PEAK 37, CGI-PDE  
DE (Fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=92283180; PubMed=1317779;  
RA Lebon T.R., Kasuya J., Paxton R.J., Belfrage P., Hockman S.,  
RA Manganiello V.C., Fujita Yamaguchi Y.;  
RT "Purification and characterization of guanosine 3',5'-monophosphate-  
RT inhibited low K(m) adenosine 3',5'-monophosphate phosphodiesterase  
RT from human placental cytosolic fractions.";  
RL Endocrinology 130:3265-3274(1992).  
FT NON-TER 1 1  
FT NON-TER 10 10  
SQ SEQUENCE 10 AA; 1272 MW; C80C440B5449C046 CRC64;  
Query Match 100.0%; Score 28; DB 4; Length 10;  
Best Local Similarity 100.0%; Pred. No. 3.1;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 NYEMK 5  
Db 6 NYEMK 10  
RESULT 2  
ID 091859 PRELIMINARY; PRT; 124 AA.  
AC 091859;

DT 01-NOV-1998 (TREMblrel. 08, Created)  
 DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE Nonstructural protein 2.  
 GN NS2.  
 OS Bovine respiratory syncytial virus.  
 OC Viruses; ssRNA negative-strand viruses; Mononegavirales;  
 OC Paramyxoviridae; Pneumovirinae; Pneumovirus.  
 OX NCBI\_TaxID=11246;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=88LUI95;  
 RA Larsen L.E., Gottschalk E., Blixenkronne-Møller M.;  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF054664; AAC3667.1; -  
 DR InterPro: IPR004336; RSV\_NS2.  
 DR Pfam: PF03113; RSV\_NS2; 1.  
 SQ SEQUENCE 124 AA; 14677 MW; 20311D79D915EB5 CRC64;

Query Match 100.0%; Score 28; DB 12; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYEMK 5  
 |||||  
 DB 64 NYEMK 68

## RESULT 3

Q9YWM9 PRELIMINARY; PRT; 124 AA.  
 AC Q9YWM9;  
 DT 01-MAY-1999 (TREMblrel. 10, Created)  
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE Nonstructural protein 2.  
 GN NS2.  
 OS Bovine respiratory syncytial virus.  
 OC Viruses; ssRNA negative-strand viruses; Mononegavirales;  
 OC Paramyxoviridae; Pneumovirinae; Pneumovirus.  
 OX NCBI\_TaxID=11246;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=88CVA70;  
 RA Larsen L.E., Gottschalk E., Blixenkronne-Møller M.;  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF054665; AAC36670.1; -  
 DR InterPro: IPR004336; RSV\_NS2.  
 DR Pfam: PF03113; RSV\_NS2; 1.  
 SQ SEQUENCE 124 AA; 14645 MW; 20311D7B2ABE9BA0 CRC64;

Query Match 100.0%; Score 28; DB 12; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYEMK 5  
 |||||  
 DB 64 NYEMK 68

## RESULT 4

Q9YWM8 PRELIMINARY; PRT; 124 AA.  
 AC Q9YWM8;  
 DT 01-MAY-1999 (TREMblrel. 10, Created)  
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE Nonstructural protein 2.  
 GN NS2.  
 OS Bovine respiratory syncytial virus.  
 OC Viruses; ssRNA negative-strand viruses; Mononegavirales;  
 OC Paramyxoviridae; Pneumovirinae; Pneumovirus.  
 OX NCBI\_TaxID=11246;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=9304899;  
 RA Larsen L.E., Gottschalk E., Blixenkronne-Møller M.;  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF054666; AAC36673.1; -  
 DR InterPro: IPR004336; RSV\_NS2.  
 DR Pfam: PF03113; RSV\_NS2; 1.  
 SQ SEQUENCE 124 AA; 14635 MW; 13311D7B29BE9AE9 CRC64;

Query Match 100.0%; Score 28; DB 12; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYEMK 5  
 |||||  
 DB 64 NYEMK 68

## RESULT 5

Q9YWM6 PRELIMINARY; PRT; 124 AA.  
 AC Q9YWM6;  
 DT 01-MAY-1999 (TREMblrel. 10, Created)  
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE Nonstructural protein 2.  
 GN NS2.  
 OS Bovine respiratory syncytial virus.  
 OC Viruses; ssRNA negative-strand viruses; Mononegavirales;  
 OC Paramyxoviridae; Pneumovirinae; Pneumovirus.  
 OX NCBI\_TaxID=11246;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=9402020;  
 RA Larsen L.E., Gottschalk E., Blixenkronne-Møller M.;  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF054667; AAC36676.1; -  
 DR InterPro: IPR004336; RSV\_NS2.  
 DR Pfam: PF03113; RSV\_NS2; 1.  
 SQ SEQUENCE 124 AA; 14649 MW; 353112D19ABE8EB5 CRC64;

Query Match 100.0%; Score 28; DB 12; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYEMK 5  
 |||||  
 DB 64 NYEMK 68

## RESULT 6

Q9YWM4 PRELIMINARY; PRT; 124 AA.  
 AC Q9YWM4;  
 DT 01-MAY-1999 (TREMblrel. 10, Created)  
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE Nonstructural protein 2.  
 GN NS2.  
 OS Bovine respiratory syncytial virus.  
 OC Viruses; ssRNA negative-strand viruses; Mononegavirales;  
 OC Paramyxoviridae; Pneumovirinae; Pneumovirus.  
 OX NCBI\_TaxID=11246;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=9314893;  
 RA Larsen L.E., Gottschalk E., Blixenkronne-Møller M.;  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF054668; AAC36679.1; -  
 DR InterPro: IPR004336; RSV\_NS2.  
 DR Pfam: PF03113; RSV\_NS2; 1.  
 SQ SEQUENCE 124 AA; 14605 MW; 23F9095B2ABC5C65 CRC64;

Query Match 100.0%; Score 28; DB 12; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NYEMK 5  
 DB 64 NYEMK 68

## RESULT 7

O9YS29 PRELIMINARY; PRT; 124 AA.  
 AC O9YS29;  
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)  
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE Nonstructural protein 2.  
 GN NS2.  
 OS Bovine respiratory syncytial virus, and  
 OS Bovine respiratory syncytial virus (strain A51908) (BRS).  
 OC Viruses; ssRNA negative-strand viruses; Mononegavirales;  
 OC Paramyxoviridae; Pneumovirinae; Pneumovirus.  
 OX NCBI\_TaxID=11246, 11247;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-Bovine respiratory syncytial virus; STRAIN-ATUES1908;  
 RX MEDLINE-99102581; PubMed-9847328;  
 RA Buchholz U.J., Flink S., Conzelmann K.K.;  
 RT "Generation of bovine respiratory syncytial virus (BRSV) from cDNA:  
 RT the human RSV leader region acts as a functional BRSV genome  
 RT promoter.";  
 RT J. Virol. 73:251-259(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-Bovine respiratory syncytial virus; STRAIN-ATUES1908;  
 RA Buchholz U.J., Flink S., Conzelmann K.K.;  
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-Bovine respiratory syncytial virus (strain A51908) (BRS);  
 RX STRAIN-ATUES1908;  
 RC MEDLINE-21580793; PubMed-11724268;  
 RA Yunus A.S., Khattar S.K., Collins P.L., Samal S.K.;  
 RT "Rescue of bovine respiratory syncytial virus from cloned cDNA: entire  
 RT genome sequence of BRSV strain A51908.";  
 RL Virus Genes 23:157-164(2001).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-Bovine respiratory syncytial virus (strain A51908) (BRS);  
 RC STRAIN-ATUES1908;  
 RA Yunus A.S., Khattar S.K., Collins P.L., Samal S.K.;  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF295543; AAC96302.1; -;  
 DR EMBL: AF295543; AAC96302.1; -;  
 DR InterPro: IPR004336; RSV\_NS2.  
 DR Pfam: PF03113; RSV\_NS2; 1.  
 SQ SEQUENCE 124 AA; 14605 MW; D80A8317FABE9D9AC CRC64;

Query Match 100.0%; Score 28; DB 12; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NYEMK 5  
 DB 64 NYEMK 68

RESULT 8  
 ID 012405 PRELIMINARY; PRT; 124 AA.  
 AC 012405;

DT 01-JUL-1997 (TREMblrel. 04, Created)  
 DT 01-JUL-1997 (TREMblrel. 04, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE Non-structural protein 2 (NS2).  
 GN NS2.  
 OS Human respiratory syncytial virus, and  
 OS respiratory syncytial virus.  
 OC Viruses; ssRNA negative-strand viruses; Mononegavirales;  
 OC Paramyxoviridae; Pneumovirinae; Pneumovirus.  
 OX NCBI\_TaxID=11250, 12814;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S2 AND S2 TSIC;  
 RX MEDLINE-97185152; PubMed-9032893;  
 RA Tolley K.P., Marriott A.C., Simpson A., Plove D.J., Matthews D.A.,  
 RA Longhurst S.J., Evans J.E., Johnson J.L., Cane P.A., Easton A.J.,  
 RA Pringle C.R.;  
 RT "Identification of mutations contributing to the reduced virulence of  
 RT a modified strain of respiratory syncytial virus.";  
 RL Vaccine 14:1637-1646(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S2 AND S2 TSIC;  
 RA Easton A.J.;  
 RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U39662; AAC57021.1; -;  
 DR EMBL: U39662; AAC57021.1; -;  
 DR InterPro: IPR004336; RSV\_NS2.  
 DR Pfam: PF03113; RSV\_NS2; 1.  
 SQ SEQUENCE 124 AA; 14729 MW; 1922830810B9B4F1 CRC64;

Query Match 100.0%; Score 28; DB 12; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NYEMK 5  
 DB 64 NYEMK 68

## RESULT 9

O9AN06 PRELIMINARY; PRT; 133 AA.  
 ID O9AN06  
 AC O9AN06;  
 DT 01-JUN-2001 (TREMblrel. 17, Created)  
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE ID28.  
 GN ID28.  
 OS Bradyrhizobium japonicum.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Bradyrhizobium group; Bradyrhizobium.  
 OX NCBI\_TaxID=375;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-110SPC4;  
 RX MEDLINE-21101824; PubMed-1157954;  
 RA Gottfert M., Rothlisberger S., Kundig C., Beck C., Marty R.,  
 RA Hennecke H.;  
 RT "Potential symbiosis-specific genes uncovered by sequencing a 410-kb  
 RT DNA region of the Bradyrhizobium japonicum chromosome.";  
 RL J. Bacteriol. 183:1405-1412(2001).  
 DR EMBL: AF322012; AAG60707.1; -;  
 DR HSRF: P37182; ICF2.  
 DR MEROPS: M52.001; -;  
 DR InterPro: IPR000671; Hydrgn\_uptake.  
 DR InterPro: IPR003662; sub\_transporter.  
 DR Pfam: PF01750; HycI; 1.  
 DR TIGRPFAMs: TIGR00072; hydrgn\_prot; 1.  
 DR PROSITE: PS00216; SUGAR\_TRANSPORT\_1; UNKNOWN\_1.  
 SQ SEQUENCE 133 AA; 14605 MW; FD371219966FD2F7 CRC64;

Query Match 100.0%; Score 28; DB 2; Length 133;

Best Local Similarity 100.0%; Pred. No. 41;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NTEMK 5  
DB 120 NTEMK 124

## RESULT 10

089012 PRELIMINARY; PRT; 320 AA.

AC 089012; 01-NOV-1998 (TREMBlrel. 08, Created)  
DT 01-NOV-1998 (TREMBlrel. 08, last sequence update)  
DE Magea8 protein.  
GN Magea8.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.

RC TISSUE=LIVER;  
RX MEDLINE=93134295; PubMed=9333564;  
RA De Plaen E., de Backer O., Arnaud D., Bonjean B., Chomez P.,  
Martelange V., Avner P., Baldacel P., Babinet C., Hwang S.Y.,  
Knowles B., Boon T.;  
RT "A new family of mouse genes homologous to the human MAGE genes.";  
RL Genomics 55:176-184(1999).  
DR EMBL; AJ005532; CA06586.1; -;  
DR MGD; MGI:1333827; Magea8.  
DR InterPro; IPR001092; HLH\_basic.  
DR InterPro; IPR002190; MAGE.  
DR Pfam; PF01454; MAGE; 1.  
DR PROSITE; PS00038; HELIX\_LOOP\_HELIX; UNKNOWN.1.  
SQ SEQUENCE 320 AA; 36214 MW; 414DEDD99253565C CRC64;

Query Match 100.0%; Score 28; DB 11; Length 320;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NTEMK 5  
DB 110 NTEMK 114

## RESULT 11

089010 PRELIMINARY; PRT; 325 AA.

AC 089010; 01-NOV-1998 (TREMBlrel. 08, Created)  
DT 01-NOV-1998 (TREMBlrel. 08, last sequence update)  
DE Magea6 protein.  
GN Magea6.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.

RC TISSUE=LIVER;  
RX MEDLINE=93134295; PubMed=9333564;  
RA De Plaen E., de Backer O., Arnaud D., Bonjean B., Chomez P.,  
Martelange V., Avner P., Baldacel P., Babinet C., Hwang S.Y.,  
Knowles B., Boon T.;  
RT "A new family of mouse genes homologous to the human MAGE genes.";  
RL Genomics 55:176-184(1999).  
DR EMBL; AJ005530; CA06584.1; -;  
DR MGD; MGI:1333837; Magea6.  
DR InterPro; IPR001092; HLH\_basic.  
DR InterPro; IPR002190; MAGE.

DR Pfam; PF01454; MAGE; 1.  
DR PROSITE; PS00038; HELIX\_LOOP\_HELIX; UNKNOWN.1.  
SQ SEQUENCE 325 AA; 36640 MW; 4F582109EC03383A CRC64;

Query Match 100.0%; Score 28; DB 11; Length 325;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NTEMK 5  
DB 110 NTEMK 114

## RESULT 12

097FL1 PRELIMINARY; PRT; 330 AA.

AC 097FL1; 01-OCT-2001 (TREMBlrel. 18, Created)  
DT 01-OCT-2001 (TREMBlrel. 18, last sequence update)  
DE Predicted hydrolase from alpha/beta family, Y0KD.B.subtilis  
ortholog.  
GN CAC2725.  
OS Clostridium acetobutylicum.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;  
OC Clostridiales; Clostridiaceae; Clostridium.  
OX NCBI\_TaxID=1488;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;  
RX MEDLINE=21359325; PubMed=11466286;  
RA Neellling J., Breton G., Omeichenko M.V., Makarova K.S., Zeng Q.,  
Gibson R., Lee H.M., Dubois J., Qiu D., Hiltl J., Wolf Y.I.,  
Tatunov R.L., Sabate F., Doucette-Stamm L., Soucaille P., Daly M.J.,  
Bennett G.N., Koonin E.V., Smith D.R.;  
RT Genome sequence and comparative analysis of the solvent-producing  
bacterium Clostridium acetobutylicum.";  
RL J. Bacteriol. 183:4823-4838(2001).  
DR EMBL; AE007770; AAK80671.1; -;  
DR InterPro; IPR000379; Ser\_estrs\_site.  
DR Hydrolase; Complete proteome.  
SQ SEQUENCE 330 AA; 37432 MW; E0C2160806B744D6 CRC64;

Query Match 100.0%; Score 28; DB 16; Length 330;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NTEMK 5  
DB 209 NTEMK 213

## RESULT 13

P91317 PRELIMINARY; PRT; 358 AA.

AC P91317; 01-MAY-1997 (TREMBlrel. 03, Created)  
DT 01-MAY-1997 (TREMBlrel. 03, last sequence update)  
DE F53E10.5 protein.  
GN F53E10.5.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,  
Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,

RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,  
 RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,  
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
 RA Watson A., Weinstock L., Wilkinson-Sprat J., Wohlman P.,  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 elegans.";  
 RL Nature 368:32-38(1994).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Beck C., Wamsley P.;  
 RT "The sequence of C. elegans cosmid F53E10.";  
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
 RN (3)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Waterston R.;  
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U88177; AAB42291.1; -;  
 SQ SEQUENCE 358 AA; 41249 MW; 288D3CA725629352 CRC64;

Query Match 100.0%; Score 28; DB 5; Length 358;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NYEMK 5  
 DB 47 NYEMK 51

## RESULT 14

O92FC7 PRELIMINARY; PRT; 420 AA.  
 AC O92FC7;  
 DT 01-MAY-1999 (TREMBlrel. 10, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Factor essential for methicillin resistance FEMa.  
 GN FEMa.  
 OS Staphylococcus haemolyticus.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
 OC Staphylococcus.  
 OX NCBI\_TaxID=1283;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC2970;  
 RA Vannuffel P., Heusterpreute M., Bouyer M., Philippe M., Gala J.-L.;  
 RT "Molecular characterization of fema from Staphylococcus hominis,  
 RT Staphylococcus saprophyticus and Staphylococcus haemolyticus and fema-  
 RT based discrimination of staphylococcal species.";  
 RL Res. Microbiol. 0:0-0(1998).  
 DR EMBL: AF099962; AAC69631.1; -;  
 DR InterPro: IPR003447; Meth\_resist.  
 DR Pfam: PF02388; FemAB; 1.  
 SQ SEQUENCE 420 AA; 49582 MW; 4492B0C6598F60C2 CRC64;

Query Match 100.0%; Score 28; DB 2; Length 420;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NYEMK 5  
 DB 29 NYEMK 33

## RESULT 15

O9S6V4 PRELIMINARY; PRT; 420 AA.  
 AC O9S6V4;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE FEMa.  
 GN FEMa.  
 OS Staphylococcus haemolyticus.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
 OC Staphylococcus.  
 OX NCBI\_TaxID=1283;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Hoskins J., Alborn J.T.W.E., Unal S., Flokowsch J.E., Greaney M.,  
 RA Skatrud P.L.;  
 RT "Cloning and Characterization of fema and femb from Staphylococcus  
 RT epidermidis and Staphylococcus haemolyticus.";  
 RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U23711; AAD22132.1; -;  
 DR InterPro: IPR003447; Meth\_resist.  
 DR Pfam: PF02388; FemAB; 1.  
 SQ SEQUENCE 420 AA; 49559 MW; 1C003811D2C59C1E CRC64;

Query Match 100.0%; Score 28; DB 2; Length 420;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NYEMK 5  
 DB 29 NYEMK 33

Search completed: July 11, 2003, 11:59:07  
 Job time : 4.18295 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: July 11, 2003, 11:42:16 ; Search time 5.73805 Seconds  
(without alignments)  
232.223 Million cell updates/sec

Title: US-10-091-442-9  
Perfect score: 50  
Sequence: 1 AVAMHGERK 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues  
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A.Geneseq-101002:\*

1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	50	100.0	380	15	AA48379
2	50	100.0	380	20	AA108254
3	50	100.0	380	21	AA24142
4	50	100.0	380	22	AA83075
5	44	88.0	10	15	AA57115
6	41	82.0	380	20	AA08255
7	41	82.0	380	21	AA24150
8	41	82.0	380	22	AA64286
9	41	82.0	380	22	AA83076
10	40	80.0	131	23	ABP34239

11	37	74.0	368	20	AA08256
12	37	74.0	368	21	AA24151
13	37	74.0	368	22	AA83077
14	36	72.0	118	22	ABG11324
15	34	68.0	67	20	AA11676
16	34	68.0	82	20	AA13595
17	34	68.0	82	21	AA12159
18	34	68.0	83	20	AAW78170
19	34	68.0	221	34	ABG11323
20	33	66.0	226	20	AA04707
21	33	66.0	223	22	AA59373
22	33	66.0	309	22	AA678271
23	33	66.0	337	20	AA34993
24	33	66.0	700	22	AAW23827
25	33	66.0	700	22	AA38831
26	33	66.0	717	22	AA40617
27	33	66.0	793	13	AAW20743
28	33	66.0	793	17	AAW02282
29	33	66.0	793	20	AA04695
30	33	66.0	1232	22	ABG63685
31	32	64.0	65	22	AA678029
32	32	64.0	136	23	ABG90000
33	32	64.0	235	21	AA020091
34	32	64.0	235	23	AA097023
35	32	64.0	236	20	AA04700
36	32	64.0	254	22	AA59372
37	32	64.0	299	20	AA89198
38	32	64.0	310	22	AA678270
39	32	64.0	474	22	ABG58359
40	32	64.0	581	22	ABG63013
41	32	64.0	643	22	ABG22551
42	32	64.0	708	22	AA893073
43	32	64.0	798	15	AA57283
44	32	64.0	802	13	AAW20744
45	32	64.0	802	17	AAW02283

## ALIGNMENTS

RESULT 1	AA48379	standard; Protein; 380 AA.
XX	AA48379;	
AC	AA48379;	
XX		
DT	16-AUG-1994	(first entry)
XX		
DE	Human megakaryocyte differentiation factor.	
XX		
KW	Human megakaryocyte differentiation factor; MDF; thrombopoietin;	
KW	hematopoietic stimulating factor; thrombocytopoiesis; platelet;	
KW	bone marrow transplantation; cancer chemotherapy.	
XX		
OS	Homo sapiens.	
XX		
PN	EP583884-A.	
XX		
PD	23-FEB-1994.	
XX		
PF	19-JUL-1993;	93BP-0305654.
XX		
PR	17-JUL-1992;	92JP-0212305.
PR	04-MAR-1993;	93JP-0067339.
XX		
PA	(SUNR ) SUNTORY LTD.	
PA	(TSUJ ) TSUJIMOTO M.	
XX		
PI	Ishida N, Iwasa F, Kurihara T, Miura K, Nakazato H;	
PI	Tsujiimoto M, Tsurooka N, Yamaguchi N, Yamachi K;	
XX		
DR	WPI; 1994-058782/08.	
DR	N-PSDB; AA056670.	

Mouse megalin prote
Mouse megalin prote
Mouse megalin prote
Novel human diagno
Human 5' EST sec
Extended human sec
Hydrophobic domain
Human secreted pro
Novel human diagno
Mouse RPRa antio
Human protein tyro
Human PTP-epsilon-
Chlamydia pneumoni
Human EST encoded
Human polypeptide
Human polypeptide
Murine receptor-ty
Murine receptor-ty
Mouse receptor-ty
Drosophila melanog
Piscirickettsia sa
Human polypeptide
Protein encoded by
Cattle enteropetl
Human RPRa amino
Human protein tyro
Aspartate-specific
Human PTP-alpha-DI
Drosophila melanog
Drosophila melanog
Novel human diagno
Human protein sega
Bovine enterokinas
Human receptor-ty
Human receptor-ty

XX New megakaryocyte differentiation factor - isolated from human  
 PT epidermoid carcinoma cells, used to treat conditions involving a  
 PT decrease in platelets  
 XX  
 PS Claim 7; Page 30-32; 47pp; English.  
 XX  
 CC Human MDF can be isolated from a culture of human epidermoid  
 CC carcinoma A431 cells in protein-free medium. The MDF stimulates  
 CC differentiation of megakaryocytes from myeloid cells in the presence  
 CC of IL-3. The MDF acts in vivo as a thrombopoietin making it useful  
 CC for treatment of diseases involving a decrease in platelet number  
 CC (esp. thrombocytopenia) such as occurs in bone marrow  
 CC transplantation and in chemotherapy. MDF has mol.wt. 55-57kD by  
 CC SDS-PAGE and exhibits an isoelectric point of 6.0-7.0.  
 XX  
 SQ Sequence 380 AA:  
 Query Match 100.0%; Score 50; DB 15; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 0.06;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 204 AVAMMHQERK 213

QY 1 AVAMMHQERK 10  
 |||||  
 DB 204 AVAMMHQERK 213

RESULT 2  
 AAY08254  
 ID AAY08254 standard; Protein; 380 AA.  
 XX  
 AC AAY08254;  
 XX  
 DT 14-JUL-1999 (first entry)  
 XX  
 DE Human megsin protein.  
 XX  
 KW Megsin; mesangial cell; treatment; diagnosis; disease; IgA nephropathy;  
 KW human; rat; murine.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO915652-A1.  
 XX  
 PD 01-APR-1999.  
 XX  
 PF 22-SEP-1998; 98WO-JP04269.  
 XX  
 PR 22-SEP-1997; 97JP-0275302.  
 XX  
 PA (KURO/) KUROKAWA K.  
 PA (MIYA/) MIYATA T.  
 XX  
 PI Miyata T;  
 XX  
 DR WPI; 1999-276983/23.  
 DR N-PSDB; AAX56712.  
 XX  
 PT Megsin protein expressed specifically in mesangial cells  
 PS Claim 1; Page 62-64; 100pp; Japanese.  
 XX  
 CC This invention describes the isolation of novel megsin nucleic acid and  
 CC proteins from human, rat and mouse tissue. This protein is expressed  
 CC specifically in mesangial cells. The products of the invention are  
 CC useful for the treatment and diagnosis of diseases involving mesangial  
 CC cells, such as IgA nephropathy.  
 XX  
 SQ Sequence 380 AA:  
 Query Match 100.0%; Score 50; DB 20; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 0.06;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AVAMMHQERK 10  
 |||||  
 DB 204 AVAMMHQERK 213

RESULT 3  
 AAB24142  
 ID AAB24142 standard; Protein; 380 AA.  
 XX  
 AC AAB24142;  
 XX  
 DT 30-JAN-2001 (first entry)  
 XX  
 DE Human megsin protein sequence SEQ ID NO:2.  
 XX  
 KW Megsin; mesangium-predominant gene; serpin regulated; nephropathy;  
 KW IgA; immunoglobulin A; detection; renal function; renal disorder;  
 KW diagnosis; biological sample; blood; urine.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200057189-A1.  
 XX  
 PD 28-SEP-2000.  
 XX  
 PF 17-MAR-2000; 2000WO-JP01646.  
 XX  
 PR 19-MAR-1999; 99JP-0075305.  
 PR 28-OCT-1999; 99JP-030623.  
 XX  
 PA (KURO/) KUROKAWA K.  
 PA (FUSO) FUSO PHARM IND LTD.  
 PA (MIYA/) MIYATA T.  
 XX  
 PI Miyata T;  
 XX  
 DR WPI; 2000-611642/58.  
 DR N-PSDB; AAA99294.  
 XX  
 PT Evaluating renal function comprises assaying megsin protein in  
 PT biological sample -  
 XX  
 PS Example 2; Page 66-69; 93pp; Japanese.  
 XX  
 CC The present invention describes a method for evaluating renal function.  
 CC The method comprises assaying megsin protein in biological sample. Also  
 CC described are: (1) use of a anti-megsin protein antibody for diagnosing  
 CC renal function; and (2) a kit for detecting megsin protein comprising:  
 CC (a) anti-megsin protein antibody attached to solid magnetic particles;  
 CC (b) direct or indirect fixing for the antibody to the particles; and  
 CC (c) a magnet. The process is useful for evaluating renal function and  
 CC diagnosing renal disorders by assaying megsin protein in biological  
 CC samples (preferably urine or blood). The process is reproducible and  
 CC gives accurate results. The present sequence represents the human megsin  
 CC protein, which is given in the exemplification of the present invention.  
 XX  
 SQ Sequence 380 AA:  
 Query Match 100.0%; Score 50; DB 21; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 0.06;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AVAMMHQERK 10  
 |||||  
 DB 204 AVAMMHQERK 213

RESULT 4  
 AAB83075  
 ID AAB83075 standard; Protein; 380 AA.  
 XX  
 AC AAB83075;



XX 10-JUL-2001 (first entry)  
 DT XX  
 DE Human megsin protein.  
 XX  
 KW Human; megsin; mesangial cell proliferative nephritis; nephrotropic;  
 KM transgenic mouse; glomerular disease; animal model; drug screening.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200124628-A1.  
 XX  
 PD 12-APR-2001.  
 XX  
 PF 06-OCT-2000; 2000WO-JP06988.  
 XX  
 PR 06-OCT-1999; 99JP-0285736.  
 XX  
 PA (KURO/) KUROKAWA K.  
 PA (MIYA/) MIYATA T.  
 XX  
 PI Miyata T;  
 XX  
 DR WPI: 2001-300136/31.  
 DR N-PSDB; AAF82438.  
 XX  
 PT Mouse model for mesangial cell proliferative nephritis for development  
 PT and screening of new treatments -  
 PS Example 4; Page 44-46; 62pp; Japanese.  
 XX  
 CC The present sequence is human megsin. The human megsin coding  
 CC sequence may be introduced into a mouse to produce an animal model of  
 CC mesangial cell proliferative nephritis. The symptoms include  
 CC enlargement of the mesangial base region, sedimentation of an immune  
 CC complex and an increase in mesangial cells. The animal model is useful  
 CC for analysing the pathology of chronic glomerular diseases and for  
 CC screening compositions for prevention and treatment of the diseases.  
 CC Highly uniform models can be made easily and in large numbers using  
 CC this method.  
 CC  
 SQ Sequence 380 AA;  
 XX  
 QY  
 DB 1 AVAMMHQERK 10  
 204 AVAMMHQERK 213  
 RESULT 5  
 AAR57115  
 ID AAR57115 standard; peptide; 10 AA.  
 XX  
 AC AAR57115;  
 XX  
 DT 16-AUG-1994 (first entry)  
 XX  
 DE Human megakaryocyte differentiation factor peptide 9.  
 XX  
 KW Human megakaryocyte differentiation factor; MPF; thrombopoietin;  
 KW haematopoietic stimulating factor; thrombocytopoiesis; platelet;  
 KW bone marrow transplantation; cancer chemotherapy.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 8 /note- "not determined"  
 XX  
 PN EP583884-A.

XX 23-FEB-1994.  
 PD XX  
 XX 19-JUL-1993; 93EP-0305654.  
 FE XX  
 PR 17-JUL-1992; 92JP-0212305.  
 PR 04-MAR-1993; 93JP-0067339.  
 XX  
 PA (SUNR) SUNTORY LTD.  
 PA (TSUJ/) TSUJIMOTO M.  
 XX  
 PI Ishida N, Iwasa F, Kurihara T, Miura K, Nakazato H;  
 PI Tsujimoto M, Tsuruoka N, Yamaguchi N, Yamachi K;  
 XX  
 DR WPI: 1994-058782/08.  
 XX  
 PT New megakaryocyte differentiation factor - isolated from human  
 PT epidermoid carcinoma cells, used to treat conditions involving a  
 PT decrease in platelets  
 PS Claim 1; Page 21; 47pp; English.  
 XX  
 CC Human MDF (see AAR48379) can be isolated from a culture of human  
 CC epidermoid carcinoma A431 cells in protein-free medium. The MDF  
 CC stimulates differentiation of megakaryocytes from myeloid cells  
 CC in the presence of IL-3. The MDF acts in vivo as a thrombopoietin  
 CC making it useful for treatment of diseases involving a decrease  
 CC in platelet number (esp. thrombocytopaenia) such as occurs in bone  
 CC marrow transplantation and in chemotherapy. MDF has mol.wt.  
 CC 55-57KD by SDS-PAGE, exhibits an isoelectric point of 6.0-7.0 and  
 CC contains an amino acid sequence comprising at least one of the  
 CC sequences AAR57107-R57115.  
 CC  
 SQ Sequence 10 AA;  
 XX  
 QY  
 DB 1 AVAMMHQERK 10  
 1 AVAMMHQXK 10  
 RESULT 6  
 AAY08255  
 ID AAY08255 standard; Protein; 380 AA.  
 XX  
 AC AAY08255;  
 XX  
 DT 14-JUL-1999 (first entry)  
 XX  
 DE Rat megsin protein.  
 XX  
 KW Megsin; mesangial cell; treatment; diagnosis; disease; Iga nephropathy;  
 KW human; rat; murine.  
 XX  
 OS Rattus rattus.  
 XX  
 PN WO9915652-A1.  
 XX  
 PD 01-APR-1999.  
 XX  
 PF 22-SEP-1998; 98WO-JP04269.  
 PF 22-SEP-1997; 97JP-0275302.  
 PR  
 XX  
 PA (KURO/) KUROKAWA K.  
 PA (MIYA/) MIYATA T.  
 XX  
 PI Miyata T;  
 XX  
 DR WPI: 1999-276983/23.

DR N-PSDB; AAX56712.  
 XX  
 PT Mesgin protein expressed specifically in mesangial cells  
 XX  
 PS Claim 1; Page 69-72; 100pp; Japanese.  
 XX  
 CC This invention describes the isolation of novel mesgin nucleic acid and  
 CC proteins from human, rat and mouse tissue. This protein is expressed  
 CC specifically in mesangial cells. The products of the invention are  
 CC useful for the treatment and diagnosis of diseases involving mesangial  
 CC cells, such as IGA nephropathy.  
 XX  
 SQ Sequence 380 AA;  
 QY  
 Db 1 AVAMMHQERK 10  
 204 AVNMHQR 213  
 Query Match 82.0%; Score 41; DB 20; Length 380;  
 Best Local Similarity 80.0%; Pred. No. 3.8;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 RESULT 7  
 AAB24150  
 ID AAB24150 standard; Protein: 380 AA.  
 AC AAB24150;  
 DT 30-JAN-2001 (first entry)  
 DE Rat mesgin protein sequence SEQ ID NO:19.  
 KW Mesgin; mesangium-predominant gene; serpin regulated; nephropathy;  
 KM IGA; immunoglobulin A; detection; renal function; renal disorder;  
 XX diagnosis; biological sample; blood; urine.  
 OS Rattus norvegicus.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 51 /note- "unspecified"  
 FT Misc-difference 94 /note- "unspecified"  
 FT  
 PN WO200057189-A1.  
 PD 28-SEP-2000.  
 PF 17-MAR-2000; 2000WO-JP01646.  
 PR 19-MAR-1999; 99JP-0075305.  
 PR 28-OCT-1999; 99JP-0306623.  
 XX  
 PA (KUROO) KUROKAWA K.  
 PA (FUSO) FUSO PHARM IND LTD.  
 PA (MIYA) MIYATA T.  
 PI Miyata T;  
 PI  
 DR WPI: 2000-611642/58.  
 DR N-PSDB; AAC55238.  
 XX  
 PT Evaluating renal function comprises assaying mesgin protein in  
 PT biological sample -  
 XX  
 PS Example 2; Page 81-84; 93pp; Japanese.  
 XX  
 CC The present invention describes a method for evaluating renal function.  
 CC The method comprises assaying mesgin protein in biological sample. Also  
 CC described are: (1) use of a anti-mesgin protein antibody for diagnosing  
 CC renal function; and (2) a kit for detecting mesgin protein comprising:  
 CC (a) anti-mesgin protein antibody attached to solid magnetic particles;

CC (b) direct or indirect fixing for the antibody to the particles; and  
 CC (c) a magnet. The process is useful for evaluating renal function and  
 CC diagnosing renal disorders by assaying mesgin protein in biological  
 CC samples (preferably urine or blood). The process is reproducible and  
 CC gives accurate results. The present sequence represents the rat mesgin  
 CC protein, which is given in the exemplification of the present invention.  
 XX  
 SQ Sequence 380 AA;  
 QY  
 Db 1 AVAMMHQERK 10  
 204 AVNMHQR 213  
 Query Match 82.0%; Score 41; DB 21; Length 380;  
 Best Local Similarity 80.0%; Pred. No. 3.8;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 RESULT 8  
 AAG64286  
 ID AAG64286 standard; Protein: 380 AA.  
 AC AAG64286;  
 DT 21-SEP-2001 (first entry)  
 DE Rat mesgin protein.  
 DE  
 XX Rat; mesgin; renal mesangial cell; mesangium proliferative nephritis.  
 XX  
 OS Rattus norvegicus.  
 XX  
 PN WO200148019-A1.  
 PD 05-JUL-2001.  
 PF 26-DEC-2000; 2000WO-JP09251.  
 PR 28-DEC-1999; 99JP-0373677.  
 XX  
 PA (KUROO) KUROKAWA K.  
 PA (MIYA) MIYATA T.  
 PI Miyata T;  
 PI  
 DR WPI: 2001-425651/45.  
 DR N-PSDB; AAG48181.  
 XX  
 PT New antibody recognizing a partial sequence of rat mesgin protein for  
 PT diagnosis of mesangium proliferative nephritis -  
 XX  
 PS Disclosure; Page 54-56; 63pp; Japanese.  
 XX  
 CC The present invention relates to a novel antibody which recognises a  
 CC peptide consisting of residues 341-354 of rat mesgin protein. The present  
 CC sequence is the protein sequence for rat mesgin, which was used in the  
 CC present invention. Mesgin is highly expressed in renal mesangial cells  
 CC and its level is elevated in mesangium proliferative nephritis. Assay of  
 CC the serum or urine level using the antibody is therefore indicative of  
 CC this type of disorder.  
 XX  
 SQ Sequence 380 AA;  
 QY  
 Db 1 AVAMMHQERK 10  
 204 AVNMHQR 213  
 Query Match 82.0%; Score 41; DB 22; Length 380;  
 Best Local Similarity 80.0%; Pred. No. 3.8;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 RESULT 9

AA	B83076
ID	AAB83076 standard; Protein: 380 AA.
AC	
XX	AAB83076:
DT	10-JUL-2001 (first entry)
XX	
DE	Rat mesgln protein.
XX	
KW	Rat: mesgln; mesangial cell proliferative nephritis; nephrotropic;
XX	transgenic mouse; glomerular disease; animal model; drug screening.
OS	Rattus norvegicus.
PN	WO200124628-A1.
PD	12-APR-2001.
PF	06-OCT-2000; 2000MO-JP06988.
PR	06-OCT-1999; 99JP-0285736.
PA	(KUROO/) KUROKAWA K.
PI	(MIYA/) MIYATA T.
DR	Miyata T:
DR	WPJ: 2001-300136/31.
DR	N-PSDB: AAF82439.
PT	Mouse model for mesangial cell proliferative nephritis for development
PT	and screening of new treatments -
PS	Disclosure: Page 48-50; 62pp; Japanese.
CC	The present sequence is rat mesgln. The human mesgln coding
CC	sequence may be introduced into a mouse to produce an animal model of
CC	mesangial cell proliferative nephritis. The symptoms include
CC	enlargement of the mesangial base region, sedimentation of an immune
CC	complex and an increase in mesangial cells. The animal model is useful
CC	for analysing the pathology of chronic glomerular diseases and for
CC	screening compositions for prevention and treatment of the diseases.
CC	Highly uniform models can be made easily and in large numbers using
CC	this method.
SO	Sequence 380 AA;
Query Match	82.0%; Score 41; DB 22; Length 380;
Best Local Similarity	80.0%; Pred. No. 3.8;
Matches	8; Conservative 1; Mismatches 1; Indels 0; Gaps 0.
OY	1 AVAMMHQERR 10     Db 204 AVNMHQRER 213
RESULT 10	
ID	ABP34239
XX	ABP34239 standard. Protein: 131 AA.
AC	
XX	ABP34239:
DT	08-JUL-2002 (first entry)
DE	Human glycoprotein-like ORF3212 protein, SEQ ID NO:6424.
XX	
KW	Human: ORF: open reading frame: ORFX: drug screening; diagnosis;
KW	disease monitoring; cytokine; cell proliferation; cell differentiation;
KW	immune modulation; haematopoiesis regulation; tissue growth;
KW	angiogenesis; activin; inhibin; chemotactic; chemokinetic; hemostatic;
KW	thrombolytic; tumour inhibition; bodily characteristic; fertility;
KW	behaviour; cancer; proliferative disorder; neurological disorder;
KW	cardiovascular disease; immune system disorder; organ transplantation;

KW tissue growth disorders; tissue regeneration disorder; diabetes mellitus;  
 KW hyothyroidism; cholesterol ester storage disease; infection; vulnerability;  
 KW vasotropic; antipruritic; antidiabetic; cytostatic; nootropic;  
 KW neuroprotective; antihypertensive; antihypertensive; antihypertensive;  
 KW candidate; hypotensive; antihypertensive; antihypertensive; immunomodulator;  
 KW dermatological; analgesic; vitruicide; antibacterial; fungicide.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200190366-A2.  
 XX  
 PD 29-NOV-2001.  
 XX  
 PF 24-MAY-2001; 2001WO-US17076.  
 XX  
 PR 24-MAY-2000; 2000US-206690P.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Leach MD, Shinkets RA;  
 XX  
 DR WPI: 2002-106200/14.  
 DR N-PSDB; ABN78265.  
 XX  
 PT Novel human polypeptides and polynucleotides useful for diagnosing,  
 PT preventing and treating cardiovascular disease, neurodegenerative,  
 PT hyperproliferative disorders and disorders related to organ  
 PT transplantation -  
 XX  
 PS Claim 10; Page 1854; 2508pp; English.  
 XX  
 XX Sequences ABP31028-ABP35561 represent 4534 novel human proteins  
 CC designated ORF (open reading frame) 1-4534, and sequences ABN75054-  
 CC ABN79587 represent cDNAs encoding them. The invention also encompasses  
 CC polypeptides at least 80% identical to the ORF-ORF4534 (collectively  
 CC referred to as ORFX) proteins, polynucleotides at least 85% identical to  
 CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX  
 CC polynucleotides, the recombinant production of ORFX proteins, antibodies  
 CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and  
 CC polypeptides, methods of screening for modulators of ORFX expression or  
 CC activity, and methods of screening individuals for a predisposition to an  
 CC ORFX-associated disorder. The ORFX proteins of the invention have a wide  
 CC range of biological activities, such as cytokine, cell proliferation,  
 CC cell differentiation, immune modulation, haematopoiesis regulation,  
 CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/  
 CC chemokinetic activity, haemostatic activity, thrombolytic activity,  
 CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,  
 CC and antineoplastic activity, and may also be involved in the determination  
 CC of bodily characteristics, fertility and behaviour. ORFX proteins,  
 CC nucleic acids and antibodies may be used in the treatment of cancers,  
 CC other proliferative disorders such as psoriasis and benign tumours,  
 CC neurological disorders such as epilepsy and Alzheimer's disease,  
 CC cardiovascular diseases, immune system disorders, disorders related to  
 CC organ transplantation, disorders of tissue growth and regeneration,  
 CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester  
 CC storage disease, and infectious diseases caused by viral, bacterial,  
 CC fungal and other pathogens. ORFX nucleic acids may also be used as a  
 CC source of primers and probes, in the detection of ORFX genomic sequences  
 CC or transcripts, in the identification and cloning of homologous  
 CC sequences, in genetic diagnosis, and in forensic biology. The ORFX  
 CC nucleic acids may additionally be used to produce transgenic animals  
 CC which may be useful for studying the function and/or activity of ORFX  
 CC protein, and in drug screening. The ORFX proteins may also be used as  
 CC immunogens to generate specific antibodies, which are useful in the  
 CC diagnosis, treatment and monitoring of ORFX-associated diseases.  
 XX  
 XX Sequence 131 AA:  
 SO  
 Query Match 80.0%; Score 40; DB 23; Length 131;  
 Best Local Similarity 88.9%; Pred. No. 2;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 1 AVAMMHQER 9

Db 90 AVAMHQR 98

## RESULT 11

ID AAY08256 standard; Protein; 368 AA.

AC AAY08256;

DT 14-JUL-1999 (first entry)

DE Mouse megalin protein.

KW Megalin; mesangial cell; treatment; diagnosis; disease; Iga nephropathy;  
human; rat; murine.

OS Mus musculus.

FT Key Location/Qualifiers  
1..368  
Protein /note="partial sequence"

PN WO915652-A1.

PD 01-APR-1999.

PF 22-SEP-1998; 98WO-JP04269.

PR 22-SEP-1997; 97JP-0275302.

PA (KURO/) KUROKAWA K.  
(MIYA/) MIYATA T.

PI Miyata T;

DR WPI; 1999-276983/23.

DR N-PSDB; AAX56714.

CC Megalin protein expressed specifically in mesangial cells

PS Claim 1; Page 76-79; 100pp; Japanese.

CC This invention describes the isolation of novel megalin nucleic acid and  
proteins from human, rat and mouse tissue. This protein is expressed  
specifically in mesangial cells. The products of the invention are  
useful for the treatment and diagnosis of diseases involving mesangial  
cells, such as Iga nephropathy.

SQ Sequence 368 AA;

Query Match 74.0%; Score 37; DB 20; Length 368;

Best Local Similarity 77.8%; Pred. No. 23;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 VAMMHOERK 10

DB 193 VAMMHOERR 201

## RESULT 12

ID AAB24151 standard; Protein; 368 AA.

AC AAB24151;

DT 30-JAN-2001 (first entry)

DE Mouse megalin protein sequence SEQ. ID NO:21.

KW Megalin; mesangium-predominant gene; serpin regulated; nephropathy;  
Iga; immunoglobulin A; detection; renal function; renal disorder;  
diagnosis; biological sample; blood; urine.

XX Mus musculus.

PN WO200057189-A1.

PD 28-SEP-2000.

PF 17-MAR-2000; 2000WO-JP01646.

PR 19-MAR-1999; 99JP-0075305.

PR 28-OCT-1999; 99JP-0306623.

PA (KURO/) KUROKAWA K.  
(FUSO ) FUSO PHARM IND. LTD.  
(MIYA/) MIYATA T.

PI Miyata T;

DR WPI; 2000-61642/58.

DR N-PSDB; AAC55239.

PT Evaluating renal function comprises assaying megalin protein in  
biological sample

PS Disclosure; Page 89-91; 93pp; Japanese.

CC The present invention describes a method for evaluating renal function.  
The method comprises assaying megalin protein in biological sample. Also  
described are: (1) use of a anti-megalin protein antibody for diagnosing  
renal function; and (2) a kit for detecting megalin protein comprising:  
(a) anti-megalin protein antibody attached to solid magnetic particles;  
(b) direct or indirect fixing for the antibody to the particles; and  
(c) a magnet. The process is useful for evaluating renal function and  
diagnosing renal disorders by assaying megalin protein in biological  
samples (preferably urine or blood). The process is reproducible and  
gives accurate results. The present sequence represents the mouse megalin  
protein, which is given in the exemplification of the present invention.

SQ Sequence 368 AA;

Query Match 74.0%; Score 37; DB 21; Length 368;

Best Local Similarity 77.8%; Pred. No. 23;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 VAMMHOERK 10

DB 193 VAMMHOERR 201

## RESULT 13

ID AAB83077 standard; Protein; 368 AA.

AC AAB83077;

DT 10-JUL-2001 (first entry)

DE Murine megalin protein.

KW Mouse; megalin; mesangial cell proliferative nephritis; nephrotic;  
transgenic mouse; glomerular disease; animal model; drug screening.

OS Mus musculus.

PN WO200124628-A1.

PD 12-APR-2001.

PF 06-OCT-2000; 2000WO-JP06988.

PR 06-OCT-1999; 99JP-0285736.

PA (KURO/) KUROKAWA K.

PA (MIVA/) MIYATA T.  
 XX  
 PI Miyata T;  
 XX  
 DR WPI: 2001-300136/31.  
 DR N-PSDB: AAF82440.  
 XX  
 PT Mouse model for mesangial cell proliferative nephritis for development  
 PT and screening of new treatments  
 XX  
 PS Disclosure: Page 52-53; 62pp: Japanese.  
 XX  
 CC The present sequence is murine mesangin protein. The human mesangin coding  
 CC sequence may be introduced into a mouse to produce an animal model of  
 CC mesangial cell proliferative nephritis. The symptoms include  
 CC enlargement of the mesangial base region, sedimentation of an immune  
 CC complex and an increase in mesangial cells. The animal model is useful  
 CC for analysing the pathology of chronic glomerular diseases and for  
 CC screening compositions for prevention and treatment of the diseases.  
 CC Highly uniform models can be made easily and in large numbers using  
 CC this method.  
 XX  
 SQ Sequence 368 AA:  
 Query Match 74.0%; Score 37; DB 22; Length 368;  
 Best Local Similarity 77.8%; Pred. No. 23;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 OY 2 VAMMHOERK 10  
 | | | | | | | | | |  
 193 VNMHHERR 201  
 DB  
 RESULT 14  
 ABG11324  
 ID ABG11324 standard; Protein: 118 AA.  
 XX  
 AC ABG11324;  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #11315.  
 XX  
 KW Human: chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PE 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEO INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI: 2001-639362/73.  
 DR N-PSDB: AAS75511.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity  
 XX  
 PS Claim 20; SEQ ID No 41683; 103pp: English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
 XX  
 SQ Sequence 118 AA:  
 Query Match 72.0%; Score 36; DB 22; Length 118;  
 Best Local Similarity 50.0%; Pred. No. 11;  
 Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 AVAMMHOERK 10  
 | | | | | | | | | |  
 49 AIAVHERR 58  
 DB  
 RESULT 15  
 AAY11676  
 ID AAY11676 standard; Protein: 67 AA.  
 XX  
 AC AAY11676;  
 XX  
 DT 16-JUN-1999 (first entry)  
 XX  
 DE Human 5' EST secreted protein SEQ ID NO:328.  
 XX  
 KW Human: secreted protein; EST; expressed sequence tag; diagnosis;  
 KW forensic; gene therapy; chromosome mapping; signal peptide;  
 KW upstream regulatory sequence; cytokine activity; cell proliferation;  
 KW differentiation; haemopoiesis regulation; tissue growth regulation;  
 KW reproductive hormone regulation; chemotactic; chemokinetic; hemostatic;  
 KW thrombolytic; anti-inflammatory; tumour inhibition.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9906439-A2.  
 XX  
 PD 11-FEB-1999.  
 XX  
 PE 31-JUL-1998; 98WO-IB01233.  
 XX  
 PR 01-AUG-1997; 97US-0904468.  
 XX  
 PA (GEST) GENSET.  
 XX  
 PI Duclert A, Dumas Milne Edwards J, Lacroix B;  
 XX  
 DR WPI: 1999-153700/13.  
 DR N-PSDB: AAX40394.  
 XX  
 PT New nucleic acids encoding human secreted proteins - obtained from  
 PT cDNA libraries derived from liver, lung, large intestine, colon,  
 PT thyroid and pancreas tissue  
 XX  
 PS Claim 27; Page 395-396; 398pp: English.  
 XX  
 CC AAX40251 to AAX40397 represent 5' expressed sequence tags (ESTs) for  
 CC human secreted proteins, and encode the proteins given in AAY11533 to

CC AAY11679, respectively. The proteins given represent the signal peptide  
 CC and an N-terminal fragment of a secreted protein. The nucleic acid  
 CC sequences can be used for producing secreted human gene products. They  
 CC can also be used to develop products for diagnosis and therapy. The  
 CC proteins obtained may have cytokine activity, cell  
 CC proliferation/differentiation activity, hematopoiesis regulating  
 CC activity, tissue growth regulating activity, reproductive hormone  
 CC regulating activity, chemotactic/chemokinetic activity, haemostatic and  
 CC thrombolytic activity, receptor/ligand activity, anti-inflammatory  
 CC activity, tumour inhibition activity or other activities. The products  
 CC can be used in forensic, gene therapy and chromosome mapping procedures.  
 CC The sequences can also be used for obtaining corresponding promoter  
 CC sequences. The nucleic acids encoding the signal peptide can be used for  
 CC directing extracellular secretion of a polypeptide or the insertion of a  
 CC polypeptide into a membrane, or importing a polypeptide into a cell.  
 XX

SQ Sequence 67 AA;

Query Match 68.08; Score 34; DB 20; Length 67;  
 Best Local Similarity 55.68; Pred. No. 16;

Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 AVAMMHQER 9  
 1:1::1:1  
 Db 18 AINVLHEER 26

Search completed: July 11, 2003, 11:54:26  
 Job time : 6.73805 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: July 11, 2003, 11:49:36 ; Search time 1.78794 Seconds  
(without alignments)  
164.563 Million cell updates/sec

Title: US-10-091-442-9

Perfect score: 50

Sequence: 1 AVAMMHQERK 10

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5a\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5b\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6a\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6b\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PT05\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfile01.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	100.0	10	2	US-08-611-977-9
2	50	100.0	380	2	US-08-472-659-34
3	50	100.0	380	2	US-08-474-661-34
4	50	100.0	380	2	US-08-611-977-34
5	44	88.0	10	2	US-08-472-659-9
6	44	88.0	10	2	US-08-474-661-9
7	33	66.0	253	2	US-08-685-992-10
8	33	66.0	253	2	US-09-144-925-10
9	33	66.0	793	1	US-08-015-985-3
10	32	64.0	30	1	US-08-287-959-14
11	32	64.0	235	3	US-08-807-151-3
12	32	64.0	235	4	US-09-478-957-3
13	32	64.0	236	1	US-08-015-985-6
14	32	64.0	254	2	US-08-685-992-9
15	32	64.0	254	2	US-09-144-925-9
16	32	64.0	299	1	US-08-773-608A-2
17	32	64.0	798	1	US-08-200-900A-2
18	32	64.0	798	5	PCT-US94-00616-2
19	32	64.0	802	1	US-08-015-985-1
20	32	64.0	1657	1	US-08-287-959-1
21	32	64.0	2071	4	US-09-415-522-6
22	31	62.0	673	4	US-09-196-387-8
23	31	62.0	949	4	US-09-196-387-10
24	31	62.0	1327	4	US-09-196-387-2
25	30	60.0	235	4	US-08-944-483-65
26	29	58.0	166	1	US-08-339-152A-24
27	29	58.0	166	2	US-08-007-999B-13

28	29	58.0	166	2	US-08-689-276A-13	Sequence 13, Appl
29	29	58.0	171	4	US-09-724-864-59	Sequence 59, Appl
30	29	58.0	238	1	US-08-289-999A-2	Sequence 2, Appl1
31	29	58.0	238	1	US-08-884-203-2	Sequence 2, Appl1
32	29	58.0	238	2	US-08-878-283-2	Sequence 2, Appl1
33	29	58.0	238	3	US-09-082-021-2	Sequence 2, Appl1
34	29	58.0	238	4	US-09-182-616-2	Sequence 2, Appl1
35	29	58.0	245	2	US-08-825-781-1	Sequence 1, Appl1
36	29	58.0	281	4	US-08-961-083-196	Sequence 196, App
37	29	58.0	511	2	US-09-073-362-1	Sequence 1, Appl1
38	29	58.0	511	2	US-09-243-920-1	Sequence 1, Appl1
39	29	58.0	539	2	US-09-057-969-4	Sequence 4, Appl1
40	29	58.0	587	2	US-08-436-664-23	Sequence 23, Appl
41	29	58.0	587	3	US-09-135-642-23	Sequence 23, Appl
42	29	58.0	587	3	US-08-394-232A-23	Sequence 23, Appl
43	29	58.0	587	5	PCT-US95-04080-23	Sequence 23, Appl
44	29	58.0	588	2	US-08-436-664-25	Sequence 25, Appl
45	29	58.0	588	3	US-09-135-642-25	Sequence 25, Appl

#### ALIGNMENTS

RESULT 1  
US-08-611-977-9  
Sequence 9, Application US/08611977  
Patent No. 5972886  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSURUOKA, No. 5972886uno  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. 5972886uhiro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAGUCHI, Kozo  
APPLICANT: YAMAGUCHI, No. 5972886omi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
City: Alexandria  
State: Virginia  
Country: United States  
Zip: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/611, 977  
FILING DATE: 06-MAR-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091, 028  
FILING DATE: 14-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-204  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:

LENGTH: 10 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-611-977-9

Query Match 100.0%; Score 50; DB 2; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.00028;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AVAMMHQERK 10  
DB 1 AVAMMHQERK 10

US-08-472-659-34  
Sequence 34, Application US/08472659  
Patent No. 5831030

## GENERAL INFORMATION:

APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IMASA, Fuyuki  
APPLICANT: TSUROOKA, No. 5831030uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. 5831030uhlro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, No. 5831030oml  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/472,659  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 536

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McCowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-248  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021

## INFORMATION FOR SEQ ID NO: 34:

SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-472-659-34

Query Match 100.0%; Score 50; DB 2; Length 380;

Best Local Similarity 100.0%; Pred. No. 0.013;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AVAMMHQERK 10  
DB 204 AVAMMHQERK 213

US-08-474-661-34  
Sequence 34, Application US/08474661  
Patent No. 5874253

## GENERAL INFORMATION:

APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IMASA, Fuyuki  
APPLICANT: TSUROOKA, No. 5874253uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. 5874253uhlro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, No. 5874253oml  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/474,661  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993

## ATTORNEY/AGENT INFORMATION:

NAME: REA, TERESA STANER  
REGISTRATION NUMBER: 30,427  
REFERENCE/DOCKET NUMBER: 001560-204  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-6620

## INFORMATION FOR SEQ ID NO: 34:

SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-474-661-34

Query Match 100.0%; Score 50; DB 2; Length 380;  
Best Local Similarity 100.0%; Pred. No. 0.013;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AVAMMHQERK 10  
DB 204 AVAMMHQERK 213

RESULT 4



US-08-611-977-34  
; Sequence 34, Application US/08611977  
; Patent No. 5972886  
; GENERAL INFORMATION:  
; APPLICANT: TSUJIMOTO, Masafumi  
; APPLICANT: IWASA, Fuyuki  
; APPLICANT: TSUROJOKA, No. 5972886uo  
; APPLICANT: NAKAZATO, Hiroshi  
; APPLICANT: MIURA, Kenju  
; APPLICANT: ISHIDA, No. 5972886uhito  
; APPLICANT: KURIHARA, Tatsuya  
; APPLICANT: YAMAGUCHI, Kozo  
; APPLICANT: YAMAGUCHI, No. 5972886oml  
; TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
; NUMBER OF SEQUENCES: 34  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: P.O. Box 1404  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/611,977  
; FILING DATE: 06-MAR-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/091,028  
; FILING DATE: 14-JUL-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 4-212305  
; FILING DATE: 17-JUL-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 6-067339  
; FILING DATE: 04-MAR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McGowan, Malcolm K.  
; REGISTRATION NUMBER: 39,300  
; REFERENCE/DOCKET NUMBER: 001560-204  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 34:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 380 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-611-977-34

Query Match 100.0%; Score 50; DB 2; Length 380;  
Best Local Similarity 100.0%; Pred. No. 0.013;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AVAMMHOERK 10  
Db 204 AVAMMHOERK 213

RESULT 5  
US-08-472-659-9  
; Sequence 9, Application US/08472659  
; Patent No. 5831030  
; GENERAL INFORMATION:  
; APPLICANT: TSUJIMOTO, Masafumi  
; APPLICANT: IWASA, Fuyuki  
; APPLICANT: TSUROJOKA, No. 5831030uo  
; APPLICANT: NAKAZATO, Hiroshi

APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. 5831030uhito  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAGUCHI, Kozo  
APPLICANT: YAMAGUCHI, No. 5831030oml  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/472,659  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-248  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-472-659-9

Query Match 88.0%; Score 44; DB 2; Length 10;  
Best Local Similarity 90.0%; Pred. No. 0.0046;  
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AVAMMHOERK 10  
Db 1 AVAMMHOXK 10

RESULT 6  
US-08-474-661-9  
; Sequence 9, Application US/08474661  
; Patent No. 5874253  
; GENERAL INFORMATION:  
; APPLICANT: TSUJIMOTO, Masafumi  
; APPLICANT: IWASA, Fuyuki  
; APPLICANT: TSUROJOKA, No. 5874253uo  
; APPLICANT: NAKAZATO, Hiroshi  
; APPLICANT: MIURA, Kenju  
; APPLICANT: ISHIDA, No. 5874253uhito  
; APPLICANT: KURIHARA, Tatsuya  
; APPLICANT: YAMAGUCHI, Kozo  
; APPLICANT: YAMAGUCHI, No. 5874253oml  
; TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
; NUMBER OF SEQUENCES: 34

;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Burns, Doane, Swecker & Mathis  
;; STREET: George Mason Bldg., Washington & Prince Sts.  
;; CITY: Alexandria  
;; STATE: Virginia  
;; COUNTRY: United States  
;; ZIP: 22313-1404  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;;  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/474, 661  
;; FILING DATE: 07-JUN-1995  
;; CLASSIFICATION: 435  
;; PRIORITY APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/091,028  
;; FILING DATE: 14-JUL-1993  
;; APPLICATION NUMBER: JP 4-212305  
;; FILING DATE: 17-JUL-1992  
;; PRIORITY APPLICATION DATA:  
;; APPLICATION NUMBER: JP 5-067339  
;; FILING DATE: 04-MAR-1993  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: REA, TERESA STANER  
;; REGISTRATION NUMBER: 30,427  
;; REFERENCE/DOCKET NUMBER: 001560-204  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (703) 836-6620  
;; TELEFAX: (703) 836-6620  
;;  
;; INFORMATION FOR SEQ ID NO: 9:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 10 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;;  
;; US-08-474-661-9

Query Match 88.0%; Score 44; DB 2; Length 10;  
Best Local Similarity 90.0%; Pred. No. 0.0046; 1; Indels 0; Gaps 0;  
Matches 9; Conservative 0; Mismatches 1;

QY 1 AVAMMHQERR 10  
DB 1 AVAMMHQXRR 10

RESULT 7  
US-08-685-992-10  
; Sequence 10, Application US/08685992  
; Patent No. 5912138  
; GENERAL INFORMATION:  
; APPLICANT: Tonks, Nicholas  
; APPLICANT: Flint, Andrew J.  
; TITLE OF INVENTION: SUBSTRATE TRAPPING PROTEIN  
; TITLE OF INVENTION: TYROSINE PHOSPHATASES  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.  
; STREET: Two Millitia Drive  
; CITY: Lexington  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02173  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows 95  
; SOFTWARE: FastSeq for Windows Version 2.0b  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/685,992

;; FILING DATE: 25-JUL-1996  
;; CLASSIFICATION: 435  
;; PRIORITY APPLICATION DATA:  
;; APPLICATION NUMBER:  
;; FILING DATE:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Granahan, Patricia  
;; REGISTRATION NUMBER: 32,227  
;; REFERENCE/DOCKET NUMBER: CSHL96-03  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 781-861-6240  
;; TELEFAX: 781-861-9540  
;;  
;; INFORMATION FOR SEQ ID NO: 10:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 253 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;;  
;; US-08-685-992-10

Query Match 66.0%; Score 33; DB 2; Length 253;  
Best Local Similarity 66.7%; Pred. No. 23;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 VAMMHQERR 10  
DB 203 VAMMHQERR 211

RESULT 8  
US-09-144-925-10  
; Sequence 10, Application US/09144925  
; Patent No. 5951979  
; GENERAL INFORMATION:  
; APPLICANT: Tonks, Nicholas  
; APPLICANT: Flint, Andrew J.  
; TITLE OF INVENTION: SUBSTRATE TRAPPING PROTEIN  
; TITLE OF INVENTION: TYROSINE PHOSPHATASES  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.  
; STREET: Two Millitia Drive  
; CITY: Lexington  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02421-4799  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows 95  
; SOFTWARE: FastSeq for Windows Version 2.0b  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/144,925  
; FILING DATE:  
; CLASSIFICATION:  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: 08/685,992  
; FILING DATE: JULY 25, 1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Granahan, Patricia  
; REGISTRATION NUMBER: 32,227  
; REFERENCE/DOCKET NUMBER: CSHL96-032  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 781-861-6240  
; TELEFAX: 781-861-9540  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 253 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single

TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-144-925-10

Query Match 66.0%; Score 33; DB 2; Length 233;  
Best Local Similarity 66.7%; Pred. No. 23;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 VAMHOKR 10  
DB 203 MAMHAEOK 211

RESULT 9  
US-08-015-985-3  
Sequence 3, Application US/08015985  
Patent No. 538886  
GENERAL INFORMATION:  
APPLICANT: Schlessler, Joseph  
APPLICANT: Sap, Jan M.  
TITLE OF INVENTION: NOVEL RECEPTOR-TYPE PHOSPHOTYROSINE  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 AVENUE OF THE AMERICAS  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/015,985  
FILING DATE: 10-FEB-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7683-020  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 793 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-015-985-3

Query Match 66.0%; Score 33; DB 1; Length 793;  
Best Local Similarity 85.7%; Pred. No. 75;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 MMHOKR 10  
DB 452 MMHSERK 458

RESULT 10  
US-08-287-959-14  
Sequence 14, Application US/08287959  
Patent No. 5639651  
GENERAL INFORMATION:  
APPLICANT: Weisbach, Lawrence  
APPLICANT: Bernards, Andre  
APPLICANT: Settleman, Jeffrey  
TITLE OF INVENTION: GAP-RELATED GENE

NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: U.S.A.  
ZIP: 02110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/287,959  
FILING DATE: August 9, 1994  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Clark, Paul C.  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 00786/181001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 30 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-287-959-14

Query Match 64.0%; Score 32; DB 1; Length 30;  
Best Local Similarity 60.0%; Pred. No. 3.9;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 AVAMHOKR 10  
DB 10 SLARHOKR 19

RESULT 11  
US-08-807-151-3  
Sequence 3, Application US/08807151  
Patent No. 6043033  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Lal, Preeti  
TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: US  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/807,151  
FILING DATE: Filed Herewith  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0227 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 235 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 416132  
US-08-807-151-3

Query Match 64.0%; Score 32; DB 3; Length 235;  
Best Local Similarity 66.7%; Pred. No. 33;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 VAMHQRK 10  
:|||||  
Db 93 IAMMHLEMK 101

RESULT 12  
US-09-478-957-3  
Sequence 3, Application US/09478957  
Patent No. 6350448  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Lal, Preeti  
TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED  
TITLE OF INVENTION: PROTEASE  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: US  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/478,957  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/807,151  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0227 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 235 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 416132  
US-09-478-957-3

Query Match 64.0%; Score 32; DB 4; Length 235;

Best Local Similarity 66.7%; Pred. No. 33;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 VAMHQRK 10  
:|||||  
Db 93 IAMMHLEMK 101

RESULT 13  
US-08-015-985-6  
Sequence 6, Application US/08015985  
Patent No. 553886  
GENERAL INFORMATION:  
APPLICANT: Schlusser, Joseph  
APPLICANT: Sap, Jan M.  
TITLE OF INVENTION: NOVEL RECEPTOR-TYPE PHOSPHOTYROSINE  
TITLE OF INVENTION: PHOSPHATASE-ALPHA  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 AVENUE OF THE AMERICAS  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/015,985  
FILING DATE: 10-FEB-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7683-020  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 236 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-015-985-6

Query Match 64.0%; Score 32; DB 1; Length 236;  
Best Local Similarity 85.7%; Pred. No. 33;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 MMHQRK 10  
:|||||  
Db 197 MMHQRK 203

RESULT 14  
US-08-685-992-9  
Sequence 9, Application US/08685992  
Patent No. 5912138  
GENERAL INFORMATION:  
APPLICANT: Tonks, Nicholas  
APPLICANT: Flint, Andrew J.  
TITLE OF INVENTION: SUBSTRATE TRAPPING PROTEIN  
TITLE OF INVENTION: TYROSINE PHOSPHATASES  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.  
STREET: Two Militia Drive

CITY: Lexington  
STATE: MA  
COUNTRY: USA  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows 95  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/685,992  
FILING DATE: 25-JUL-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: CSHL96-03  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 781-861-6240  
TELEFAX: 781-861-9540  
TELEX:  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 254 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-685-992-9

Query Match 64.0%; Score 32; DB 2; Length 254;  
Best Local Similarity 85.7%; Pred. No. 36;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 4 MMHQERK 10  
DB 206 MMHTERK 212

RESULT 15  
US-09-144-925-9  
Sequence 9, Application US/09144925  
Patent No. 5951979  
GENERAL INFORMATION:  
APPLICANT: Tonks, Nicholas  
APPLICANT: Flint, Andrew J.  
TITLE OF INVENTION: SUBSTRATE TRAPPING PROTEIN  
TITLE OF INVENTION: TYROSINE PHOSPHATASES  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.  
STREET: Two Millita Drive  
CITY: Lexington  
STATE: MA  
COUNTRY: USA  
ZIP: 02421-4799  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows 95  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/144,925  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/685,992  
FILING DATE: July 25, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia

REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: CSHL96-032  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 781-861-6240  
TELEFAX: 781-861-9540  
TELEX:  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 254 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-144-925-9

Query Match 64.0%; Score 32; DB 2; Length 254;  
Best Local Similarity 85.7%; Pred. No. 36;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 4 MMHQERK 10  
DB 206 MMHTERK 212

Search completed: July 11, 2003, 12:02:12  
Job time: 2.78794 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: July 11, 2003, 11:59:17 ; Search time 2.93139 Seconds  
(without alignments)  
397.148 Million cell updates/sec

Title: US-10-091-442-9  
Perfect score: 50  
Sequence: 1 AVAMMHQERK 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 445758 seqs, 116419773 residues

Total number of hits satisfying chosen parameters: 445758

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

Published Applications-AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	50	100.0	10	9	US-10-091-442-9
2	50	100.0	10	10	US-09-140-719-9
3	50	100.0	380	9	US-10-091-442-34
4	50	100.0	380	10	US-09-140-719-34
5	33	66.0	309	10	US-09-788-626-11
6	32	64.0	235	10	US-09-988-975A-6
7	32	64.0	279	9	US-09-858-332-15
8	32	64.0	310	10	US-09-788-626-10
9	31	62.0	318	10	US-09-919-497-91
10	31	62.0	318	10	US-09-919-497-92
11	31	62.0	370	9	US-10-128-714-8591
12	31	62.0	673	10	US-09-841-835-8
13	31	62.0	718	10	US-09-815-242-10901
14	31	62.0	949	10	US-09-841-835-10
15	31	62.0	1347	10	US-09-972-115A-8
16	31	62.0	1347	9	US-09-841-835-2
17	31	62.0	54	10	US-09-864-761-47786
18	30	60.0	82	9	US-09-956-622A-38
19	30	60.0			Sequence 38, Appl

## ALIGNMENTS

20	30	60.0	377	9	US-10-165-605A-27	Sequence 27, Appl
21	30	60.0	377	10	US-09-910-430-27	Sequence 27, Appl
22	30	60.0	427	8	US-08-808-031A-40	Sequence 40, Appl
23	30	60.0	427	10	US-09-925-297-684	Sequence 684, App
24	30	60.0	1005	9	US-09-934-070-8	Sequence 8, Appl1
25	30	60.0	1019	9	US-10-157-031-267	Sequence 267, App
26	30	60.0	1019	9	US-09-776-191-64	Sequence 64, Appl
27	29	58.0	155	9	US-09-866-050A-699	Sequence 699, App
28	29	58.0	173	10	US-09-867-550-888	Sequence 888, App
29	29	58.0	261	9	US-10-156-761-8516	Sequence 8516, App
30	29	58.0	281	10	US-09-765-272-196	Sequence 196, App
31	29	58.0	285	9	US-10-156-761-12080	Sequence 12080, A
32	29	58.0	314	9	US-09-925-299-934	Sequence 934, App
33	29	58.0	314	9	US-10-106-698-4725	Sequence 4725, Ap
34	29	58.0	314	10	US-09-925-299-934	Sequence 934, App
35	29	58.0	322	10	US-09-764-853-606	Sequence 606, App
36	29	58.0	356	9	US-10-123-965B-5	Sequence 5, Appl1
37	29	58.0	427	10	US-09-815-242-13443	Sequence 13443, A
38	29	58.0	427	10	US-09-815-242-13640	Sequence 13640, A
39	29	58.0	458	10	US-09-815-242-11759	Sequence 11759, A
40	29	58.0	511	9	US-10-163-866-41	Sequence 41, Appl
41	29	58.0	511	9	US-10-163-866-42	Sequence 42, Appl
42	29	58.0	638	9	US-10-101-464A-74	Sequence 74, Appl
43	29	58.0	734	9	US-10-040-862-10463	Sequence 10463, A
44	29	58.0	773	9	US-10-112-286-2	Sequence 2, Appl1
45	28	56.0	50	9	US-10-106-698-7062	Sequence 7062, Ap

RESULT 1  
US-10-091-442-9  
Sequence 9, Application US/10091442  
Patent No. US20020164711A1  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
IMASA, Eiyuki  
TSUKAZONO, No. US20020164711A1uo  
NAKAZATO, Hiroshi  
MURA, Kenju  
ISHIDA, No. US20020164711A1uhiro  
KURIHARA, Tatsuya  
YAMACHI, Kozo  
YAMAGUCHI, No. US20020164711A1omi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/091,442  
FILING DATE: 07-Mar-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
APPLICATION NUMBER: JP 6-067339

ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 9:  
US-10-091-442-9

Query Match 100.0%; Score 50; DB 9; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.00028;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AVAMMHQERK 10  
DB 1 AVAMMHQERK 10

RESULT 2  
US-09-140-719-9  
Sequence 9, Application US/09140719  
Patent No. US20010026931A1  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSUROOKA, No. US20010026931A1uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. US20010026931A1uhro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, No. US20010026931A1om1  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.

REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-140-719-9

Query Match 100.0%; Score 50; DB 10; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.00028;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AVAMMHQERK 10  
DB 1 AVAMMHQERK 10

RESULT 3  
US-10-091-442-34  
Sequence 34, Application US/10091442  
Patent No. US20020164711A1  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSUROOKA, No. US20020164711A1uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. US20020164711A1uhro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, No. US20020164711A1om1  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/091,442  
FILING DATE: 07-MAR-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021



INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 34:  
US-10-091-442-34

Query Match 100.0%; Score 50; DB 9; Length 380;  
Best Local Similarity 100.0%; Pred. No. 0.019;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AVAMHGERK 10  
DB 204 AVAMHGERK 213

RESULT 4  
US-09-140-719-34  
Sequence 34, Application US/09140719  
Patent No. US20010026931A1  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IMASA, Fuyuki  
APPLICANT: TSUROJOKA, No. US20010026931A1uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MURA, Kenju  
APPLICANT: ISHIDA, No. US20010026931A1uhltro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, No. US20010026931A1om1  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
APPLICATION DATA:  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid

TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-140-719-34

Query Match 100.0%; Score 50; DB 10; Length 380;  
Best Local Similarity 100.0%; Pred. No. 0.019;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AVAMHGERK 10  
DB 204 AVAMHGERK 213

RESULT 5  
US-09-788-626-11  
Sequence 11, Application US/09788626  
Patent No. US20020009762A1  
GENERAL INFORMATION:  
APPLICANT: Flint, Andrew J.  
APPLICANT: COOL, Deborah E.  
TITLE OF INVENTION: IMPROVED ASSAY FOR PROTEIN TYROSINE  
FILE OF INVENTION: PHOSPHATES  
FILE REFERENCE: 200125,401  
CURRENT APPLICATION NUMBER: US/09/788,626  
CURRENT FILING DATE: 2001-02-13  
NUMBER OF SEQ ID NOS: 40  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 11  
LENGTH: 309  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-788-626-11

Query Match 66.0%; Score 33; DB 10; Length 309;  
Best Local Similarity 66.7%; Pred. No. 51;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 VAMHGERK 10  
DB 203 VAMHGERK 211

RESULT 6  
US-09-988-975A-6  
Sequence 6, Application US/09988975A  
Patent No. US20020119531A1  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Lal, Preeti G.  
TITLE OF INVENTION: PROSTATE-ASSOCIATED PROTEASE ANTIBODY  
FILE REFERENCE: PP-0227-2 CIP  
CURRENT APPLICATION NUMBER: US/09/988,975A  
CURRENT FILING DATE: 2001-11-19  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: PERL Program  
SEQ ID NO 6  
LENGTH: 235  
TYPE: PRT  
ORGANISM: Homo sapiens  
PUBLICATION INFORMATION:  
DATABASE ACCESSION NUMBER: Genbank ID No. US20020119531A1 g416132  
US-09-988-975A-6

Query Match 64.0%; Score 32; DB 10; Length 235;  
Best Local Similarity 66.7%; Pred. No. 60;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 VAMHGERK 10  
DB 93 VAMHGERK 101

RESULT 7

US-09-858-332-15  
; Sequence 15, Application US/09858332  
; Patent No. US20020164718A1  
; GENERAL INFORMATION:  
; APPLICANT: Tchaqa, Grigory S.  
; APPLICANT: Jokhadze, George  
; TITLE OF INVENTION: Metal Ion Affinity Tags and Methods for  
; FILE REFERENCE: CLON056CIP  
; CURRENT APPLICATION NUMBER: US/09/858,332  
; CURRENT FILING DATE: 2002-07-02  
; PRIOR APPLICATION NUMBER: 09/404,017  
; PRIOR FILING DATE: 1999-09-23  
; PRIOR APPLICATION NUMBER: 60/101,867  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15  
; LENGTH: 279  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: synthetic peptide  
US-09-858-332-15

Query Match 64.0%; Score 32; DB 9; Length 279;  
Best Local Similarity 66.7%; Pred. No. 73;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 VAMMHOERK 10  
Db 122 IAMMHEMK 130

RESULT 8  
US-09-788-626-10  
; Sequence 10, Application US/09788626  
; Patent No. US2002009762A1  
; GENERAL INFORMATION:  
; APPLICANT: Flint, Deborah J.  
; APPLICANT: Cool, Deborah E.  
; TITLE OF INVENTION: IMPROVED ASSAY FOR PROTEIN TYROSINE  
; FILE REFERENCE: 200125.401  
; CURRENT APPLICATION NUMBER: US/09/788,626  
; CURRENT FILING DATE: 2001-02-13  
; NUMBER OF SEQ ID NOS: 40  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10  
; LENGTH: 310  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-788-626-10

Query Match 64.0%; Score 32; DB 10; Length 310;  
Best Local Similarity 85.7%; Pred. No. 82;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 MMHOERK 10  
Db 206 MMHTEK 212

RESULT 9  
US-09-919-497-91  
; Sequence 91, Application US/09919497  
; Patent No. US2002010662A1  
; GENERAL INFORMATION:  
; APPLICANT: Mutter, George L.  
; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER  
; FILE REFERENCE: B0801/7225  
; CURRENT APPLICATION NUMBER: US/09/919,497  
; CURRENT FILING DATE: 2001-07-31

; PRIOR APPLICATION NUMBER: US 60/221,735  
; PRIOR FILING DATE: 2000-07-31  
; NUMBER OF SEQ ID NOS: 100  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 91  
; LENGTH: 318  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-919-497-91

Query Match 62.0%; Score 31; DB 10; Length 318;  
Best Local Similarity 62.5%; Pred. No. 14e+02;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 3 AMMHOERK 10  
Db 190 ALIHKEK 197

RESULT 10  
US-09-919-497-92  
; Sequence 92, Application US/09919497  
; Patent No. US2002010662A1  
; GENERAL INFORMATION:  
; APPLICANT: Mutter, George L.  
; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER  
; FILE REFERENCE: B0801/7225  
; CURRENT APPLICATION NUMBER: US/09/919,497  
; CURRENT FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: US 60/221,735  
; PRIOR FILING DATE: 2000-07-31  
; NUMBER OF SEQ ID NOS: 100  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 92  
; LENGTH: 318  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-919-497-92

Query Match 62.0%; Score 31; DB 10; Length 318;  
Best Local Similarity 62.5%; Pred. No. 14e+02;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 3 AMMHOERK 10  
Db 190 ALIHKEK 197

RESULT 11  
US-10-128-714-8591  
; Sequence 8591, Application US/10128714  
; Publication No. US20030119013A1  
; GENERAL INFORMATION:  
; APPLICANT: Jiang, Bo  
; APPLICANT: Hu, Weng  
; APPLICANT: Tishkoff, Daniel  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Eroshkin, Alexey M  
; APPLICANT: Lemieux, Sebastien M  
; TITLE OF INVENTION: Identification of Essential Genes in *Aspergillus fumigatus* and  
; FILE REFERENCE: 10182-018-999  
; CURRENT APPLICATION NUMBER: US/10/128,714  
; CURRENT FILING DATE: 2002-04-23  
; PRIOR APPLICATION NUMBER: US 60/285,697  
; PRIOR FILING DATE: 2001-04-23  
; PRIOR APPLICATION NUMBER: US 60/287,066  
; PRIOR FILING DATE: 2001-04-27  
; PRIOR APPLICATION NUMBER: US 60/295,890  
; PRIOR FILING DATE: 2001-06-05  
; PRIOR APPLICATION NUMBER: US 60/303,899  
; PRIOR FILING DATE: 2001-07-09  
; PRIOR APPLICATION NUMBER: US 60/316,362

;; PRIOR FILING DATE: 2001-08-31  
;; NUMBER OF SEQ ID NOS: 8603  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO: 8591  
;; LENGTH: 370  
;; TYPE: PRF  
;; ORGANISM: Aspergillus fumigatus  
US-10-128-714-8591

Query Match 62.0%; Score 31; DB 9; Length 370;  
Best Local Similarity 75.0%; Pred. No. 1.6e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 VAMMOER 9  
DB 311 VAMRHKR 318

RESULT 12  
US-10-128-714-3591  
;; Sequence 3591, Application US/10128714  
;; Publication No. US20030119013A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Jiang, Bo  
;; APPLICANT: Hu, Meng  
;; APPLICANT: Tishkoff, Daniel  
;; APPLICANT: Zamudio, Carlos  
;; APPLICANT: Ecosmkin, Alexey M  
;; APPLICANT: Lemieux, Sebastien M  
;; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and  
;; FILE REFERENCE: 10182-018-999  
;; CURRENT APPLICATION NUMBER: US/10/128,714  
;; CURRENT FILING DATE: 2002-04-23  
;; PRIOR APPLICATION NUMBER: US 60/285,697  
;; PRIOR FILING DATE: 2001-04-23  
;; PRIOR APPLICATION NUMBER: US 60/287,066  
;; PRIOR FILING DATE: 2001-04-27  
;; PRIOR APPLICATION NUMBER: US 60/295,890  
;; PRIOR FILING DATE: 2001-06-05  
;; PRIOR APPLICATION NUMBER: US 60/303,899  
;; PRIOR FILING DATE: 2001-07-09  
;; PRIOR APPLICATION NUMBER: US 60/316,362  
;; PRIOR FILING DATE: 2001-08-31  
;; NUMBER OF SEQ ID NOS: 8603  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO: 3591  
;; LENGTH: 445  
;; TYPE: PRF  
;; ORGANISM: Aspergillus fumigatus  
US-10-128-714-3591

Query Match 62.0%; Score 31; DB 9; Length 445;  
Best Local Similarity 75.0%; Pred. No. 2e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 VAMMOER 9  
DB 386 VAMRHKR 393

RESULT 13  
US-09-841-835-8  
;; Sequence 8, Application US/09841835  
;; Patent No. US20020076795A1  
;; GENERAL INFORMATION:  
;; APPLICANT: de Lange, Titia  
;; APPLICANT: Smith, Susan  
;; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS  
;; NUMBER OF SEQUENCES: 12  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Klauber & Jackson

STREET: 411 Hackensack Avenue, 4th Floor  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/841,835  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/196,387  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-230 CIP1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-487-5800  
TELEFAX: 201-343-1684  
TELEX: 133521

INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 673 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-841-835-8

Query Match 62.0%; Score 31; DB 10; Length 673;  
Best Local Similarity 60.0%; Pred. No. 3.3e+02;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 AVAMMOER 10  
DB 529 AVASLHPRK 538

RESULT 14  
US-09-815-242-10901  
;; Sequence 10901, Application US/09815242  
;; Patent No. US20020061569A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Haselbeck, Robert  
;; APPLICANT: Ohlsen, Karl L.  
;; APPLICANT: Zyskind, Judith W.  
;; APPLICANT: Wall, Daniel  
;; APPLICANT: Trawick, John D.  
;; APPLICANT: Carr, Grant J.  
;; APPLICANT: Yamamoto, Robert T.  
;; APPLICANT: Xu, H. Howard  
;; TITLE OF INVENTION: Identification of Essential Genes in  
;; FILE REFERENCE: ELITRA.01A  
;; CURRENT APPLICATION NUMBER: US/09/815,242  
;; CURRENT FILING DATE: 2001-03-21  
;; PRIOR APPLICATION NUMBER: 60/191,078  
;; PRIOR FILING DATE: 2000-03-21  
;; PRIOR APPLICATION NUMBER: 60/206,848  
;; PRIOR FILING DATE: 2000-05-23  
;; PRIOR APPLICATION NUMBER: 60/207,727  
;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR APPLICATION NUMBER: 60/242,578  
;; PRIOR FILING DATE: 2000-10-23  
;; PRIOR APPLICATION NUMBER: 60/253,625  
;; PRIOR FILING DATE: 2000-11-27  
;; PRIOR APPLICATION NUMBER: 60/257,931  
;; PRIOR FILING DATE: 2000-12-22  
;; PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16  
 NUMBER OF SEQ ID NOS: 1410  
 SOFTWARE: FASTSEQ for Windows Version 4.0  
 SEQ ID NO 10901  
 LENGTH: 718  
 TYPE: PRT  
 ORGANISM: Enterococcus faecalis  
 US-09-815-242-10901

Job time : 3.93139 secs

Query Match 62.0%; Score 31; DB 10; Length 718;  
 Best Local Similarity 66.7%; Pred. No. 3.5e+02;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 AVAMMOER 9  
 ||| |:  
 Db 122 AVAMMOER 130

RESULT 15  
 US-09-841-835-10  
 Sequence 10, Application US/09841835  
 Patent No. US20020076795A1  
 GENERAL INFORMATION:  
 APPLICANT: de Lange, Titia  
 APPLICANT: Smith, Susan  
 TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS  
 TITLE OF INVENTION: OF USE THEREOF  
 NUMBER OF SEQUENCES: 12  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Klauber & Jackson  
 STREET: 411 Hackensack Avenue, 4th Floor  
 CITY: Hackensack  
 STATE: New Jersey  
 COUNTRY: USA  
 ZIP: 07601  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/841,835  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 09/196,387  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Jackson Esq., David A.  
 REGISTRATION NUMBER: 26,742  
 REFERENCE/DOCKET NUMBER: 600-1-230 CIP1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 201-487-5800  
 TELEFAX: 201-343-1684  
 TELEX: 133521  
 INFORMATION FOR SEQ ID NO: 10:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 949 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-09-841-835-10

Query Match 62.0%; Score 31; DB 10; Length 949;  
 Best Local Similarity 60.0%; Pred. No. 4.8e+02;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 AVAMMOER 10  
 ||| |:  
 Db 529 AVASLHPRK 538

Search completed: July 11, 2003, 12:37:35

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: July 11, 2003, 11:45:06 ; Search time 2.01663 Seconds  
(without alignments)  
476.708 Million cell updates/sec

Title: US-10-091-442-9  
Perfect score: 50  
Sequence: 1 AVAMHGERK 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	70.0	374	2 T34435	hypothetical prote
2	35	70.0	578	2 H86177	hypothetical prote
3	35	70.0	602	2 A84744	hypothetical prote
4	35	70.0	1561	2 T00248	zinc finger protei
5	33	66.0	275	2 H70234	replicative DNA he
6	33	66.0	281	2 AD2052	hypothetical prote
7	33	66.0	327	1 E35270	6-phosphotricokin
8	33	66.0	337	1 D72084	holliday junction
9	33	66.0	337	2 D86539	holliday junction
10	33	66.0	337	2 A81586	holliday junction
11	33	66.0	395	2 PC1143	finger protein zfc
12	33	66.0	700	1 S12053	protein-tyrosine-p
13	33	66.0	796	1 JC1285	protein-tyrosine-p
14	33	66.0	829	1 A47373	protein-tyrosine-p
15	32	64.0	229	2 C86399	protein F17L21.16
16	32	64.0	252	2 A13408	glycoprotein/poly
17	32	64.0	298	2 E70912	hypothetical prote
18	32	64.0	384	2 H64612	hypothetical prote
19	32	64.0	401	2 A47014	probable arylsulfa
20	32	64.0	431	1 JX0364	antithrombin III
21	32	64.0	474	2 A40721	neuroblast prolif
22	32	64.0	486	2 F83390	hypothetical prote
23	32	64.0	496	2 JC5170	probable methylgl
24	32	64.0	656	1 B44954	fumate reductase
25	32	64.0	802	1 A36065	protein-tyrosine-p
26	32	64.0	1035	1 A43090	enteropeptidase (E
27	32	64.0	1126	2 S49208	transmembrane prot
28	32	64.0	1657	2 A54854	Ras GTPase activat
29	31	62.0	138	1 TTR08B	thyrotropin beta c

30	31	62.0	138	1 TTPGB	thyrotropin beta c
31	31	62.0	148	2 T24628	hypothetical prote
32	31	62.0	216	2 AE0648	nitrate/nitrite re
33	31	62.0	217	2 B83944	resolvase BH254 I
34	31	62.0	217	2 F83502	hypothetical prote
35	31	62.0	236	2 A82287	phosphate transpor
36	31	62.0	253	1 G69109	conserved hypothet
37	31	62.0	301	2 T36520	hypothetical prote
38	31	62.0	318	1 KIHUR2	ribose-phosphate d
39	31	62.0	318	1 KIHUR2	ribose-phosphate d
40	31	62.0	318	1 KIHUR1	ribose-phosphate d
41	31	62.0	318	1 KIHUR1	ribose-phosphate d
42	31	62.0	318	1 KIHUR3	ribose-phosphate d
43	31	62.0	334	2 C69148	drpp-glucose 4,6-d
44	31	62.0	397	2 D64690	signal-transducing
45	31	62.0	506	2 D71461	hypothetical prote

## ALIGNMENTS

RESULT 1  
T34435  
hypothetical protein K11H12.9 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
R:Bradsheaw, H.  
submitted to the EMBL Data Library, February 1997  
A:Description: The sequence of C. elegans cosmid K11H12.  
A:Reference number: 221526  
A:Accession: T34435  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-374 <BRA>  
A:Cross-references: EMBL:U88168; PIDN:AAC24403.1; GSPDB:GN00022; CESP:K11H12.9  
A:Experimental source: strain Bristol N2; clone K11H12  
C:Genetics:  
A:Gene: CESP:K11H12.9  
A:Map position: 4  
A:Introns: 65/2; 218/3; 254/1; 301/1; 355/1

Query Match  
Best Local Similarity 70.0%; Score 35; DB 2; Length 374;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AVAMHMQ 7  
|||||||  
DB 206 AVAMHMQ 212

RESULT 2  
H86177  
hypothetical protein [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: H86177  
R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,  
ansen, N.F.; Hughes, B.; Hulzar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Malt, R.; Marzia  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tello  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MIMD:21016719; PMID:11130712  
A:Accession: H86177  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-578 <STO>  
A:Cross-references: GB:AE005172; NID:g2494130; PIDN:AA80639.1; GSPDB:GN00141

```
C:Genetics:
A:Map position: 1

Query Match
Best Local Similarity 70.0%; Score 35; DB 2; Length 578;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 AVAMMHQERK 10
    |||:| |||
Db 29 AVAMVHSERK 38

RESULT 3
A:Accession: A84744
hypothetical protein At2g33320 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: A84744
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
Guss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: A84744
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-602 <STO>
A:Cross-references: GB:AE002093; NID:g2459415; PIDN:AAB80650.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g33320
A:Map position: 2

Query Match
Best Local Similarity 70.0%; Score 35; DB 2; Length 602;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 AVAMMHQERK 10
    |||:| |||
Db 31 AVAMVHSERK 40

RESULT 4
T00248
zinc finger protein wizl - mouse
C:Species: Mus musculus (house mouse)
C:Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 05-Nov-1999
C:Accession: T00248
R:Matsumoto, K.; Ishii, N.; Yoshida, S.; Shiosaka, S.; Wanaka, A.; Tohyama, M.
submitted to the EMBL Data Library, March 1998
A:Description: Molecular cloning and distinct developmental expression pattern of splice
A:Reference number: Z14130
A:Accession: T00248
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1561 <MAT>
A:Cross-references: EMBL:AB012265; NID:d1227740; PIDN:BAA32790.1; PID:d1033756
A:Experimental source: brain
C:Genetics:
A:Gene: wizl

Query Match
Best Local Similarity 70.0%; Score 35; DB 2; Length 1561;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 AVAMMHQERK 10
    |||:| |||
Db 604 AVAMKHEERK 613

RESULT 5
H70234
replicative DNA helicase homolog - Lyme disease spirochete plasmid g/lp28-2

C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
C:Accession: H70234
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh
son, D.; Peterson, J.; Kervavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vu
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943; PMID:9403685
A:Accession: H70234
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-275 <KLE>
A:Cross-references: GB:AE00786; NID:g2690008; PIDN:AAC6061.1; PID:g2690018; TIGR:BB
A:Experimental source: strain B31
C:Genetics:
A:Genome: plasmid

Query Match
Best Local Similarity 66.0%; Score 33; DB 2; Length 275;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 2 VAMMHQERK 10
    |||:| |||
Db 223 VIFLHQERK 231

RESULT 6
AD2052
hypothetical protein all1970 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AD2052
R:Kanevo, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigu
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AD2052
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-281 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAW3669.1; PID:g17131060; GSPDB:GN00179
C:Genetics:
A:Gene: all1970

Query Match
Best Local Similarity 77.8%; Score 33; DB 2; Length 281;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 VAMMHQERK 10
    |||:| |||
Db 174 VAKHHQERK 182

RESULT 7
E35270
6-phosphofructokinase (EC 2.7.1.11) - Spiroplasma citri
N:Alternate names: phosphofructokinase 1; phosphohexokinase
C:Species: Spiroplasma citri
C:Date: 17-Aug-1990 #sequence_revision 14-Jul-1994 #text_change 07-Dec-1999
C:Accession: E35270
R:Chevalier, C.; Salllard, C.; Bove, J.M.
J. Bacteriol. 172, 2693-2703, 1990
A:Title: Organization and nucleotide sequences of the Spiroplasma citri genes for rib
in.
A:Reference number: A35270; MUID:90236934; PMID:2139649
A:Accession: E35270
A:Molecule type: DNA
A:Residues: 1-327 <CHE>
```

A:Cross-references: GB:M31161; NID:9152884; PID:9152889  
A:Note: the authors translated the codon GCT for residue 195 as Ile  
C:Genetics:  
A:Genetic code: SGC3  
C:Superfamily: 6-phosphofructokinase; 6-phosphofructokinase 1 homology  
C:Keywords: ATP; phosphotransferase  
F:5'-280/Domain: 6-phosphofructokinase 1 homology <6PFI>

Query Match 66.0%; Score 33; DB 1; Length 327;  
Best Local Similarity 66.7%; Pred. No. 28;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 VAMMHOERK 10  
||:|:|:|  
Db 206 VAMLHODKK 214

RESULT 8  
D72084  
holliday junction helicase - Chlamydomonas pneumoniae (strain CML029)

C:Species: Chlamydomonas pneumoniae, Chlamydia pneumoniae  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 05-May-2000  
C:Accession: D72084  
R:Kallman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;  
Nature Genet. 21, 385-389, 1999  
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
A:Reference number: A72000; MUID:99206606; PMID:10192388  
A:Accession: D72084  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-337 <RNA>  
A:Cross-references: GB:AE001623; GB:AE001363; NID:94376662; PIDN:AMD18533.1; PID:9437666  
A:Experimental source: strain CML029  
C:Genetics:  
A:Gene: ruvB  
C:Superfamily: ruvB protein

Query Match 66.0%; Score 33; DB 2; Length 337;  
Best Local Similarity 55.6%; Pred. No. 29;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 2 VAMMHOERK 10  
||:|:|:|  
Db 5 VAVLHODKK 13

RESULT 9

D86539  
holliday junction helicase [imported] - Chlamydomonas pneumoniae (strain J138)

C:Species: Chlamydomonas pneumoniae, Chlamydia pneumoniae  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001  
C:Accession: D86539  
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; IS  
Nucleic Acids Res. 28, 2311-2314, 2000  
A:Title: Comparison of whole genome sequences of Chlamydia pneumoniae J138.  
A:Reference number: A86491; MUID:20330349; PMID:10871362  
A:Accession: D86539  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-337 <STO>  
A:Cross-references: GB:BA000008; NID:98978762; PIDN:BA98598.1; GSPDB:GN00142  
A:Experimental source: strain J138  
C:Genetics:  
A:Gene: ruvB  
C:Superfamily: ruvB protein

Query Match 66.0%; Score 33; DB 2; Length 337;  
Best Local Similarity 55.6%; Pred. No. 29;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 2 VAMMHOERK 10  
||:|:|:|  
Db 5 VAVLHODKK 13

RESULT 10

A81586

Holliday junction DNA helicase ruvB CP0365 [imported] - Chlamydomonas pneumoniae (str  
C:Species: Chlamydomonas pneumoniae, Chlamydia pneumoniae  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-May-2000  
C:Accession: A81586

R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heldelberg, J.F.; White, O.; Hicke  
C.; Dodson, R.; Gwin, M.; Nelson, W.; Deboy, R.; Koloney, J.; McClarty, G.; Salze  
Nucleic Acids Res. 28, 1397-1406, 2000

A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39  
A:Reference number: A81500; MUID:20150255; PMID:10684935

A:Accession: A81586  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-337 <REA>  
A:Cross-references: GB:AE002198; GB:AE002161; NID:97189279; PIDN:AAF38214.1; PID:9718  
A:Experimental source: strain AR39, HL cells  
C:Genetics:  
A:Gene: CP0365  
C:Superfamily: ruvB protein

Query Match 66.0%; Score 33; DB 2; Length 337;  
Best Local Similarity 55.6%; Pred. No. 29;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 2 VAMMHOERK 10  
||:|:|:|  
Db 5 VAVLHODKK 13

RESULT 11

PC1143

finger protein zfc - American alligator (fragment)

C:Species: Alligator mississippiensis (American alligator)

C:Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 01-Dec-2000  
C:Accession: PC1143

R:Valley, E.M.A.; Mueller, U.; Ferguson, M.W.J.; Sharpe, P.T.  
Gene 119, 221-228, 1992

A:Title: Cloning and expression analysis of two zfy-related zinc finger genes from Al  
A:Reference number: PC1143; MUID:93012990; PMID:1398103

A:Accession: PC1143  
A:Molecule type: DNA

A:Residues: 1-395 <VAL>

A:Cross-references: EMBL:X61714; NID:962415; PIDN:CAA43863.1; PID:962416

C:Genetics:  
A:Gene: zfc

C:Superfamily: zinc finger protein zfp-36; LIM metal-binding repeat homology  
C:Keywords: DNA binding; zinc finger

Query Match 66.0%; Score 33; DB 2; Length 395;  
Best Local Similarity 75.0%; Pred. No. 34;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 3 AMMHOERK 10  
|:|:|:|  
Db 214 ALMHODKK 221

RESULT 12

S12053

protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type epsilon precursor - human

C:Species: Homo sapiens (man)

C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: S12053

R:Krueger, N.X.; Streuli, M.; Saito, H.  
EMBO J. 9, 3241-3252, 1990

A:Title: Structural diversity and evolution of human receptor-like protein tyrosine p  
A:Reference number: S12049; MUID:91006018; PMID:2170109

A:Accession: S12053  
A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-700 <KR>  
A:Cross-references: GB:X54134; NID:935791; PIDN:CAA38069.1; PID:935792  
C:Genetics:  
A:Gene: GDB:PTPR2  
A:Cross-references: GDB:131385; OMIM:600926  
A:Map position: 10q26-10q26  
C:Function:  
A:Description: catalyzes the hydrolysis of peptidyl-phosphotyrosine to release phosphate  
A:Superfamily: protein-tyrosine-phosphatase, receptor type alpha; leukocyte common antigen  
C:Keywords: phosphoprotein; phosphoric monoester hydrolase; receptor; transmembrane protein  
F:1-19/Domain: signal sequence #status predicted <Sig>  
F:20-700/Product: protein-tyrosine-phosphatase, receptor type epsilon #status predicted  
F:47-63/Domain: transmembrane #status predicted <TM>  
F:78-698/Domain: leukocyte common antigen cytosolic domain homology <LAC>  
F:159-383/Domain: protein-tyrosine-phosphatase homology <PRP1>  
F:335/Active site: Cys (phosphocysteine intermediate) #status predicted  
F:341/Binding site: substrate phosphatase (Arg) #status predicted  
F:630/Active site: Cys (phosphocysteine intermediate) #status predicted  
F:636/Binding site: substrate phosphatase (Arg) #status predicted

Query Match 66.0%; Score 33; DB 1; Length 700;  
Best Local Similarity 66.7%; Pred. No. 63;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 VAMMOERK 10  
DB 352 MAMMAEOK 360

RESULT 13  
JCI285  
protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type alpha precursor - rat  
N:Alternate names: protein-tyrosine-phosphatase LRP  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 10-Sep-1999 #sequence-revision 10-Sep-1999 #text-change 10-Sep-1999  
C:Accession: JCI285; S23253; JH0450  
R:Moriyama, T.; Fujiwara, Y.; Imai, E.; Takenaka, M.; Kawanishi, S.; Inoue, T.; Noguchi, T.  
Biochem. Biophys. Res. Commun. 188, 34-39, 1992  
A:Title: cDNA cloning of rat LRP, a receptor like protein tyrosine phosphatase, and evid  
A:Reference number: JCI285; MUID:93038682; PMID:1417854  
A:Accession: JCI285  
A:Molecule type: mRNA  
A:Residues: 1-796 <MOR>  
A:Experimental source: kidney  
A:Note: The authors translated the codon TCC for residue 788 as Ala  
R:Hashimoto, N.; Zhang, W. R.; Goldstein, B. J.  
Biochem. J. 284, 569-576, 1992  
A:Title: Insulin receptor and epidermal growth factor receptor dephosphorylation by three  
A:Reference number: S23126; MUID:92287069; PMID:1599438  
A:Accession: S23253  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 254-267, 'I', 269-354, 'T', 356-501 <HSA>  
R:Zhang, W. R.; Goldstein, B. J.  
Biochem. Biophys. Res. Commun. 178, 1291-1297, 1991  
A:Title: Identification of skeletal muscle protein-tyrosine phosphatases by amplification  
A:Reference number: JH0450; MUID:91337074; PMID:1651716  
A:Accession: JH0450  
A:Molecule type: mRNA  
A:Residues: 324-354, 'T', 356-434 <ZHA>  
A:Experimental source: skeletal muscle, strain Sprague-Dawley  
C:Function:  
A:Description: catalyzes the hydrolysis of peptidyl-phosphotyrosine to release phosphate  
A:Note: Allows recovery from insulin stimulation by dephosphorylating insulin receptor  
A:Superfamily: protein-tyrosine-phosphatase, receptor type alpha; leukocyte common antigen  
C:Keywords: phosphoprotein; phosphoric monoester hydrolase; receptor; transmembrane protein  
F:1-19/Domain: signal sequence #status predicted <Sig>  
F:20-796/Product: protein-tyrosine-phosphatase, receptor type alpha #status predicted <M  
F:20-148/Domain: extracellular #status predicted <EXT>  
F:149-165/Domain: transmembrane #status predicted <TM>  
F:166-796/Domain: intracellular #status predicted <INT>  
F:178-794/Domain: leukocyte common antigen cytosolic domain homology <LAC>  
F:259-484/Domain: protein-tyrosine-phosphatase homology <PRP1>

F:552-774/Domain: protein-tyrosine-phosphatase homology <PRP2>  
F:436/Active site: Cys (phosphocysteine intermediate) #status predicted  
F:442/Binding site: substrate phosphatase (Arg) #status predicted  
F:726/Active site: Cys (phosphocysteine intermediate) #status predicted  
F:732/Binding site: substrate phosphatase (Arg) #status predicted

Query Match 66.0%; Score 33; DB 1; Length 796;  
Best Local Similarity 85.7%; Pred. No. 72;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 MMHSEK 10  
DB 455 MMHSEK 461

RESULT 14  
A47373  
protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type alpha precursor - mouse  
N:Alternate names: Leukocyte common antigen-related protein LRP; PTPY28; receptor pr  
C:Species: Mus musculus (house mouse)  
C:Date: 10-Sep-1999 #sequence-revision 10-Sep-1999 #text-change 21-Jan-2000  
C:Accession: A47373; B35501; A35501; P50367; S40286; A36004; D61180  
R:Nong, E. C.; Mullersman, J. E.; Thomas, M. L.  
Genomics 17, 33-38, 1993  
A:Title: Leukocyte common antigen-related phosphatase (LRP) gene structure: conservat  
A:Reference number: A47373; MUID:94010906; PMID:8406469  
A:Accession: A47373  
A:Molecule type: DNA  
A:Residues: 1-829 <MON>  
A:Cross-references: GB:L13607  
A:Note: Sequence extracted from NCBI backbone (NCBIN:137623, NCBI:137624)  
R:Matthews, R. J.; Cahill, E. D.; Thomas, M. L.  
Proc. Natl. Acad. Sci. U.S.A. 87, 4444-4448, 1990  
A:Title: Identification of an additional member of the protein-tyrosine-phosphatase f  
A:Reference number: A35501; MUID:90280391; PMID:2162042  
A:Accession: B35501  
A:Molecule type: mRNA  
A:Residues: 1-829 <MA1>  
A:Cross-references: GB:M36033; NID:9198876; PIDN:AAA39448.1; PID:9198877; GB:M33671  
A:Note: The authors translated the codon GAT for residue 30 as Tyr  
A:Accession: A35501  
A:Molecule type: mRNA  
A:Residues: 1-267, 'Y', 305-829 <MA2>  
A:Cross-references: GB:M36033; GB:M33671  
R:den Hertog, J.; Pals, C. E. G. M.; Jonk, L. J. C.; Krulder, W.  
Biochem. Biophys. Res. Commun. 184, 1241-1249, 1992  
A:Title: Differential expression of a novel murine non-receptor protein tyrosine phos  
A:Reference number: JH0609; MUID:92272714; PMID:1590786  
A:Accession: P50367  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 322-325, 'G', 327-356 <DEN>  
A:Experimental source: embryonal carcinoma cell, P19 cell  
R:Hendriks, W.; Brugman, C.; Zeeuwen, P.; Schepens, J.; Wieringa, B.  
Submitted to the EMBL Data Library, June 1993  
A:Description: Assessment of the expression levels of murine protein-tyrosine phosphat  
A:Reference number: S40280  
A:Accession: S40285  
A:Molecule type: mRNA  
A:Residues: 358-467 <HEN>  
A:Cross-references: EMBL:223054; NID:9438145; PIDN:CAA80589.1; PID:9438146  
A:Accession: S40286  
A:Molecule type: mRNA  
A:Residues: 651-756 <HE2>  
A:Cross-references: EMBL:223055; NID:9438147; PIDN:CAA80590.1; PID:9438148  
R:Sap, J.; D'Enstachio, P.; Givol, D.; Schlesinger, J.  
Proc. Natl. Acad. Sci. U.S.A. 87, 6112-6116, 1990  
A:Title: Cloning and expression of a widely expressed receptor tyrosine phosphatase.  
A:Reference number: A36004; MUID:90349565; PMID:2166945  
A:Accession: A36004  
A:Molecule type: mRNA  
A:Residues: 1-230, 'L', 232-267, 'Y', 305-410, 'S', 412-829 <SAP>  
A:Cross-references: GB:M34668



Job time : 4.11663 secs

R:YL, T.: Cleveland, J.L.; Inle, J.N.

Blood 78, 2222-2228, 1991

A:Title: Identification of novel protein tyrosine phosphatases of hematopoietic cells by

A:Reference number: A61180, PMID:92032882; PMID:1932742

A:Accession: D61180

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 358-467 &lt;YIA&gt;

C:Genetics:

A:Map position: 2

C:Function:

A:Description: catalyzes the hydrolysis of peptidyl-phosphotyrosine to release phosphate

A:Note: allows recovery from insulin stimulation by dephosphorylating insulin receptor a

C:Superfamily: protein-tyrosine-phosphatase, receptor type alpha; leukocyte common anti

C:Keywords: alternative splicing; glycoprotein; phosphoprotein; phosphoric monoester hyd

F:1-19/Domain: signal sequence #status predicted &lt;SIG&gt;

F:20-829/Product: protein-tyrosine-phosphatase, receptor type alpha #status predicted &lt;M

F:40-142/Domain: extracellular #status predicted &lt;EXT&gt;

F:143-166/Domain: transmembrane #status predicted &lt;TM&gt;

F:167-829/Domain: intracellular #status predicted &lt;INT&gt;

F:175-827/Domain: leukocyte common antigen cytosolic domain homology &lt;LAC&gt;

F:21,47,51,68,80,86,104,124/Binding site: carbohydrate (asn) (covalent) #status predicte

F:469/Active site: Cys (phosphocysteine intermediate) #status predicted

F:475/Binding site: substrate phosphate (Arg) #status predicted

F:759/Active site: Cys (phosphocysteine intermediate) #status predicted

F:765/Binding site: substrate phosphate (Arg) #status predicted

## Query Match

Best Local Similarity 66.0%; Score 33; DB 1; Length 829;

Best Local Similarity 85.7%; Pred. No. 76;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 MMHQRK 10  
||| |||  
DB 488 MMHQRK 494

## RESULT 15

C86399

protein F17L21.16 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: C86399

R:Theologas, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Huizer, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lan, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; PMID:21016719; PMID:11130712

A:Accession: C86399

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-229 &lt;RTO&gt;

A:Cross-references: GB:AEO05172; NID:g9802545; PIDN:AAF99747.1; GSPDB:GN00141

C:Genetics:

A:Gene: F17L21.16

A:Map position: 1

## Query Match

Best Local Similarity 64.0%; Score 32; DB 2; Length 229;

Best Local Similarity 60.0%; Pred. No. 32;

Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 AVAMHQRK 10  
||| ||:|  
DB 24 AVATTHQRK 33

Search completed: July 11, 2003, 12:00:44

THIS PAGE BLANK (USE ...)

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:43:41 ; Search time 0.956341 Seconds

(Without alignments)  
433.698 Million cell updates/sec

Title: US-10-091-442-9

Sequence: 1 AVAMHGERK 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	50	100.0	380	1	SPB7_HUMAN
2	33	66.0	380	1	SPB7_HUMAN
3	33	66.0	337	1	K6PF_SPIPI
4	33	66.0	700	1	RUVB_CHLPN
5	33	66.0	796	1	PTPE_HUMAN
6	33	66.0	829	1	PTRA_MOUSE
7	32	64.0	299	1	ICEL_SPOFR
8	32	64.0	327	1	KSF5_ECOLI
9	32	64.0	401	1	CHUR_BACTN
10	32	64.0	474	1	ANA_DROME
11	32	64.0	496	1	MGLA_TREPA
12	32	64.0	656	1	FRDA_WOLSU
13	32	64.0	802	1	PTRA_HUMAN
14	32	64.0	1035	1	ENTK_BOVIN
15	32	64.0	1126	1	HEM_DROME
16	32	64.0	1657	1	IOGI_HUMAN
17	32	64.0	1657	1	IOGI_MOUSE
18	31	62.0	138	1	TSBH_BOVIN
19	31	62.0	138	1	TSBH_HORSE
20	31	62.0	138	1	TSBH_PIG
21	31	62.0	138	1	TSBH_PIG
22	31	62.0	317	1	KPR2_HUMAN
23	31	62.0	317	1	KPR2_HUMAN
24	31	62.0	317	1	KPR2_RAT
25	31	62.0	317	1	KPR3_HUMAN
26	31	62.0	1327	1	TNKL_HUMAN
27	31	62.0	1503	1	MRP6_HUMAN
28	30	60.0	156	1	RNH_ZYMO
29	30	60.0	331	1	YHBU_ECOLI
30	30	60.0	340	1	ILVC_CAMEL
31	30	60.0	344	1	DHB1_MOUSE
32	30	60.0	365	1	TRUA_HUMAN
33	30	60.0	388	1	GALI_STACA

34	30	60.0	393	1	TRUA_MOUSE	Q9W56 mus musculus
35	30	60.0	422	1	SAH1_PYRAB	Q9UY5 PYROCOCCUS
36	30	60.0	427	1	RT65_MYXXA	P23071 MYXOCOCCUS
37	30	60.0	433	1	ANT3_BOVIN	P41361 bos taurus
38	30	60.0	485	1	ER24_FUSSO	Q01447 fustarium so
39	30	60.0	512	1	ER24_SEPLY	O13597 septoria ly
40	30	60.0	875	1	SYA_YERPE	Q82B8 yersinia pe
41	30	60.0	883	1	CHB_VIBRA	P13670 vibrio harv
42	30	60.0	1019	1	ENTK_HUMAN	P98073 homo sapien
43	30	60.0	1034	1	ENTK_PIG	P98074 sus scrofa
44	30	60.0	1069	1	ENTK_MOUSE	P97433 mus musculus
45	30	60.0	1169	1	SUV3_DROME	P20193 drosophila

## ALIGNMENTS

RESULT 1	SPB7_HUMAN	STANDARD	PRT	380 AA.
ID	SPB7_HUMAN	075635		
AC	15-JUN-2002 (Rel. 41, Created)			
DT	15-JUN-2002 (Rel. 41, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Megsin (TP55) (Serpin B7).			
GN	SERPINB7.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_Taxid:9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
EX	MEDLINE-97326116; PubMed-9182567;			
RA	Tsujiimoto M., Tsunoka N., Ishida N., Kurihara T., Iwasa F.,			
RA	Yamashiro K., Rogi T., Kodama S., Katsunagi N., Adachi M.,			
RA	Katayama T., Nakao M., Yamauchi K., Hashino J., Haruyama M., Miura K.,			
RA	Nakanishi T., Nakazato H., Teramura M., Mizoguchi H., Yamaguchi N.;			
RT	"Purification, cDNA cloning, and characterization of a new serpin with			
RT	megakaryocyte maturation activity."			
RL	J. Biol. Chem. 272:15373-15380(1997).			
RL	[2]			
RP	SEQUENCE FROM N.A.			
EX	TISSUE-Mesangial cells:			
RA	MEDLINE-98376492; PubMed-9710452;			
RA	Miyata T., Nangaku M., Suzuki D., Inagi R., Uragami K., Sakai H.,			
RA	Okubo K., Kurokawa K.;			
RT	"A mesangium-predominant gene, megin, is a new serpin upregulated in			
RT	IGA nephropathy."			
RL	J. Clin. Invest. 102:828-836(1998).			
CC	-I- FUNCTION: Might function as an inhibitor of lys-specific			
CC	proteases. Might influence the maturation of megakaryocytes via			
CC	its action as a serpin.			
CC	-I- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).			
CC	-I- TISSUE SPECIFICITY: Predominantly expressed in mesangial cells.			
CC	-I- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.1sb-sib.ch/announce/">http://www.1sb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@1sb-sib.ch">license@1sb-sib.ch</a> ).			
CC	-----			
DR	EMBL; D88575; BAA31232.1; -			
DR	EMBL; AF027866; AAC64506.1; -			
DR	HSSP; P05619; 1HE.			
DR	Genew; HGNC:13902; SERPINB7.			
DR	MTM; 603357; -			
DR	InterPro; IPR000215; Serpin.			
DR	Pfam; PF00079; serpin; 1.			
DR	SMART; SM00093; SERPIN; 1.			
DR	PROSITE; PS00284; SERPIN; 1.			

KW Serpin; Serine protease inhibitor.  
 FT ACT\_SITE 347 348 REACTIVE\_BOND (BY SIMILARITY).  
 SQ SEQUENCE 380 AA; 42904 MW; 9A2CDB6C63CF605 CRC64;

Query Match 100.0%; Score 50; DB 1; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 0.0025;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AVAMHQRK 10  
 |||||  
 DB 204 AVAMHQRK 213

RESULT 2  
 K6PF\_SPICI STANDARD; PRT; 327 AA.  
 ID K6PF\_SPICI  
 AC P20275;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE 6-phosphofructokinase (EC 2.7.1.11) (Phosphofructokinase)  
 DE (Phosphohexokinase).  
 GN PFKA.  
 OS Spiroplasma citri.  
 OC Bacteria; Firmicutes; Mollicutes; Entomoplasmatales;  
 OC Spiroplasmataceae; Spiroplasma.  
 OX NCBI\_TaxId=2133;  
 RN NCBI\_TaxId=2133;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 27556 / R8A2;  
 RX MEDLINE=90236934; Pubmed=2139649;  
 RA Chevallier C., Sallard C., Bove J.M.;  
 RT "Organization and nucleotide sequences of the Spiroplasma citri genes  
 for ribosomal protein S2, elongation factor Ts, spiralin,  
 phosphofructokinase, pyruvate kinase, and an unidentified protein.";  
 RT J. Bacteriol. 172:2693-2703(1990).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 27556 / R8A2;  
 RL le Dantec L.;  
 RT Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: ATP + D-fructose 6-phosphate -> ADP + D-  
 fructose 1,6-bisphosphate.  
 CC -1- PATHWAY: KEY CONTROL STEP OF GLYCOLYSIS.  
 CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE PHOSPHOFRUCTOKINASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@sdb.ch](mailto:license@sdb.ch)).  
 CC -----  
 CC EMBL; AF012877; AAB69998.1; -  
 DR PIR; E35270; E35270.  
 DR HSSP; P06998; 2PFK.  
 DR InterPro: IPR000023; pffructinase.  
 DR Pfam; PF00365; PFK; 1.  
 DR PRINTS; PR00476; PFRCTKINASE.  
 DR PRODOM; PD000707; pfprfructinase; 1.  
 DR PROSITE; PS00433; PHOSPHOFRUCTOKINASE; 1.  
 KW Kinase; Transferase; Glycolysis.  
 SQ SEQUENCE 327 AA; 35197 MW; 19D9D47E176FE2F6 CRC64;

Query Match 66.0%; Score 33; DB 1; Length 327;  
 Best Local Similarity 66.7%; Pred. No. 10;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

2 VAMHQRK 10  
 |||||

DB 206 VAMHQRK 214

RESULT 3

ID RUVB\_CHLPN STANDARD; PRT; 337 AA.  
 AC Q928F3; Q9JRV2;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Holiday junction DNA helicase ruvb.  
 DE RUVB OR CPN0390 OR CP0365.  
 GN Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OS Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxId=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CWL029;  
 RX MEDLINE=99206606; Pubmed=10192388;  
 RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";  
 RL Nat. Genet. 21:385-389(1999).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=AR39;  
 RX MEDLINE=20150255; Pubmed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Uterback T., Berry K., Bass S.,  
 RA Linher K., Weidman J., Knout H., Craven B., Bowman C., Dodson R.,  
 RA Gwin J., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,  
 RA Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis MOpn and Chlamydia  
 pneumoniae AR39.";  
 RT J. Bacteriol. 181:1397-1406(2000).  
 RL [3]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=J138;  
 RX MEDLINE=20330349; Pubmed=10871362;  
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 RA Shiba T., Ishi K., Hattori M., Kuhara S., Nakazawa T.;  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 from Japan and CWL029 from USA.";  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 CC -1- FUNCTION: THE RUVA-RUVB COMPLEX IN THE PRESENCE OF ATP RENATURES  
 CC CRICIFORM STRUCTURE IN SUPERCOILED DNA WITH PALINDROMIC SEQUENCE,  
 CC INDICATING THAT IT MAY PROMOTE STRAND EXCHANGE REACTIONS IN  
 CC HOMOLOGOUS RECOMBINATION. RUVA IS AN HELICASE THAT MEDIATES THE  
 CC HOLIDAY JUNCTION MIGRATION BY LOCALIZED DENATURATION AND  
 CC REANNEALING (BY SIMILARITY).  
 CC -1- SUBUNIT: FORMS A COMPLEX WITH RUVA (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE RUVA FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@sdb.ch](mailto:license@sdb.ch)).  
 CC -----  
 CC EMBL; AE001623; AAD18533.1; -  
 DR EMBL; AE002198; AAF38214.1; -  
 DR EMBL; AP002546; BAA98598.1; -  
 DR TIGR; CP0365; -  
 DR InterPro: IPR003593; AAA\_ATPase.  
 DR InterPro: IPR003595; AAA\_ATPase\_cent.  
 DR InterPro: IPR004605; RUVB.  
 DR Pfam; PF00004; AAA; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR TIGRFRAMS; TIGR00635; ruvb; 1.  
 KW DNA repair; SOS response; ATP-binding; DNA recombination; Helicase;  
 KW Complete proteome.

FT NE\_BIND 57 64 ATP (POTENTIAL).  
 FT CONFLICT 168 168 T -> A (IN REF. 1).  
 SO SEQUENCE 337 AA: 37398 MW: D591B593F342EA2F CRC64;

Query Match 66.0%; Score 33; DB 1; Length 337;  
 Best Local Similarity 55.6%; Pred. No. 11;  
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 2 VAMHDERK 10  
 :|||||:  
 Db 5 VAVLHDERK 13

## RESULT 4

PTRE\_HUMAN STANDARD; PRT; 700 AA.  
 ID PTRE\_HUMAN  
 AC P23469;  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Protein-tyrosine phosphatase epsilon precursor (EC 3.1.3.48) (R-PTP-epsilon).  
 GN PTPRE.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=91006018; PubMed=2170109;  
 RA Krueger N.X., Streuli M., Saito H.;  
 RT "Structural diversity and evolution of human receptor-like protein tyrosine phosphatases."  
 RL EMBO J. 9:3241-3252(1990).  
 CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O - protein tyrosine + phosphate.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: X54134; CAA38069.1; -  
 DR HSP: S12053; S12053.  
 DR HSP: P18052; IYFO.  
 DR Genew: HGNC:9669; PTPRE.  
 DR MIM: 600926; -  
 DR InterPro: IPR000387; TYR\_phosphatase.  
 DR InterPro: IPR000242; TYR\_PP.  
 DR Pfam: PF00102; Y\_phosphatase\_2.  
 DR PRINTS: PR00700; PTPPPHASE.  
 DR SMART: SM00194; PTPC; 2.  
 DR PROSITE: PS00383; TYR\_PHOSPHATASE\_1; 2.  
 DR PROSITE: PS50056; TYR\_PHOSPHATASE\_2; 2.  
 DR PROSITE: PS50055; TYR\_PHOSPHATASE\_PP; 2.  
 KW Glycoprotein; Transmembrane; Hydrolase; Phosphorylation; Repeat; Signal.  
 FT SIGNAL 1 19 POTENTIAL.  
 FT CHAIN 20 700 PROTEIN-TYROSINE PHOSPHATASE EPSILON.  
 FT DOMAIN 20 46 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 47 69 POTENTIAL.  
 FT DOMAIN 70 700 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 154 393 PROTEIN-TYROSINE PHOSPHATASE 1.  
 FT DOMAIN 394 700 PROTEIN-TYROSINE PHOSPHATASE 2.  
 FT ACT\_SITE 335 335 BY SIMILARITY.  
 FT ACT\_SITE 630 630 BY SIMILARITY.  
 FT CARBOHYD 23 23 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 30 30 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SO SEQUENCE 700 AA: 80641 MW: D096BCADCEA65708 CRC64;

Query Match 66.0%; Score 33; DB 1; Length 700;  
 Best Local Similarity 66.7%; Pred. No. 24;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 VAMHDERK 10  
 :|||||:  
 Db 352 MAMHDERK 360

## RESULT 5

PTRA\_RAT STANDARD; PRT; 796 AA.  
 ID PTRA\_RAT  
 AC Q03348;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Protein-tyrosine phosphatase alpha precursor (EC 3.1.3.48) (R-PTP-alpha).  
 GN PTPRA OR LRP.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RX MEDLINE=93038682; PubMed=1417854;  
 RA Moriyama T., Fujiwara Y., Imai E., Takenaka M., Kawanishi S., Itoe T., Noguchi T., Tanaka T., Ueda N.;  
 RT "cDNA cloning of rat LRP, a receptor like protein tyrosine phosphatase, and evidence for its gene regulation in cultured rat mesangial cells."  
 RL Biochem. Biophys. Res. Commun. 188:34-39(1992).  
 CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O - protein tyrosine + phosphate.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: L01702; AAA41983.1; -  
 DR HSP: P18052; IYFO.  
 DR InterPro: IPR000387; TYR\_phosphatase.  
 DR InterPro: IPR000242; TYR\_PP.  
 DR Pfam: PF00102; Y\_phosphatase\_2.  
 DR PRINTS: PR00700; PTPPPHASE.  
 DR SMART: SM00194; PTPC; 2.  
 DR PROSITE: PS00383; TYR\_PHOSPHATASE\_1; 2.  
 DR PROSITE: PS50056; TYR\_PHOSPHATASE\_2; 2.  
 DR PROSITE: PS50055; TYR\_PHOSPHATASE\_PP; 2.  
 KW Glycoprotein; Transmembrane; Hydrolase; Phosphorylation; Repeat; Signal.  
 FT SIGNAL 1 19 POTENTIAL.  
 FT CHAIN 20 796 PROTEIN-TYROSINE PHOSPHATASE ALPHA.  
 FT DOMAIN 20 145 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 146 169 POTENTIAL.  
 FT DOMAIN 170 796 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 234 494 PROTEIN-TYROSINE PHOSPHATASE 1.  
 FT DOMAIN 495 796 PROTEIN-TYROSINE PHOSPHATASE 2.  
 FT ACT\_SITE 436 436 BY SIMILARITY.  
 FT ACT\_SITE 726 726 BY SIMILARITY.  
 FT CARBOHYD 20 20 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 21 21 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 47 47 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 51 51 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 80 80 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 796 AA; 90260 MW; 4793796191056920 CRC64;

Query Match 66.0%; Score 33; DB 1; Length 796;  
 Best Local Similarity 85.7%; Pred. No. 28;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 MMHQERK 10  
 DB 455 MMHSERK 461  
 III III

RESULT 6  
 PTRA\_MOUSE STANDARD; PRT; 829 AA.  
 ID PTRA\_MOUSE 18052; Q61808;  
 AC 01-NOV-1990 (Rel. 16, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Protein-tyrosine phosphatase alpha precursor (EC 3.1.3.48) (R-PTP-  
 alpha) (LCA-related phosphatase).  
 GN PTRA OR LRP OR PRPA.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN (1)  
 RC SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND TISSUE SPECIFICITY.  
 RP STRAIN-C57BL/6 X DBA/2;  
 RX MEDLINE=90280391; PubMed=2162042;  
 RA Matthews R.J., Cahill E.D., Thomas M.L.;  
 RT "Identification of an additional member of the protein-tyrosine-  
 phosphatase family: evidence for alternative splicing in the tyrosine  
 phosphatase domain.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:4444-4448(1990).  
 RN (2)  
 RP SEQUENCE OF 358-467 FROM N.A.  
 RC STRAIN-BALB/c; TISSUE-Brain;  
 RX MEDLINE=95134232; PubMed=7832766;  
 RA Hendriks W., Schepens J., Brugman C., Zeeuwen P., Wieringa B.;  
 RT "A novel receptor-type protein tyrosine phosphatase with a single  
 catalytic domain is specifically expressed in mouse brain.";  
 RL Biochem. J. 305:499-504(1995).  
 RN (3)  
 RP SEQUENCE OF 651-756 FROM N.A.  
 RC STRAIN-BALB/c; TISSUE-Brain;  
 RX MEDLINE=93086603; PubMed=1454056;  
 RA Schepens J., Zeeuwen P., Wieringa B., Hendriks W.;  
 RT "Identification and typing of members of the protein-tyrosine  
 phosphatase gene family expressed in mouse brain.";  
 RL Mol. Biol. Rep. 16:241-248(1992).  
 RN (4)  
 RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 202-503.  
 RX MEDLINE=96320562; PubMed=8700232;  
 RA Bilwes A.M., den Hertog J., Hunter T., Noel J.P.;  
 RT "Structural basis for inhibition of receptor protein-tyrosine  
 phosphatase-alpha by dimerization.";  
 RL Nature 382:555-559(1996).  
 CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O -> protein  
 tyrosine + phosphate.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND  
 A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED.  
 CC -1- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----

DR EMBL; M36033; AAA39448.1; -  
 DR EMBL; M36034; AAA39449.2; -  
 DR EMBL; 223054; CAA80589.1; -  
 DR EMBL; 223055; CAA80590.1; -  
 DR PIR; A35501; A35501.  
 DR PIR; B35501; B35501.  
 DR PDB; 1YFO; 01-APR-97.  
 DR MGD; MGI:97808; Pcpa.  
 DR InterPro; IPR000387; TYR-phosphatase.  
 DR InterPro; IPR000242; TYR-PP.  
 DR Pfam; PF00102; Y-phosphatase. 2.  
 DR PRINTS; PR00700; PRTYPHTASE.  
 DR SMART; SM00194; PTPc; 2.  
 DR PROSITE; PS00383; TYR-PHOSPHATASE\_1; 2.  
 DR PROSITE; PS00056; TYR-PHOSPHATASE\_2; 2.  
 DR PROSITE; PS50055; TYR-PHOSPHATASE\_PTP; 2.  
 KW Glycoprotein; Transmembrane; Hydroxylase; Phosphorylation; Signal;  
 KW Repeat; Alternative splicing; 3D-structure.  
 FT SIGNAL 1 19  
 FT CHAIN 20 829  
 FT DOMAIN 20 142  
 FT TRANSMEM 143 166  
 FT DOMAIN 167 829  
 FT DOMAIN 232 527  
 FT DOMAIN 528 829  
 FT ACT\_SITE 469 469  
 FT ACT\_SITE 759 759  
 FT CARBOHYD 21 21  
 FT CARBOHYD 47 47  
 FT CARBOHYD 51 51  
 FT CARBOHYD 68 68  
 FT CARBOHYD 80 80  
 FT CARBOHYD 86 86  
 FT CARBOHYD 104 104  
 FT CARBOHYD 124 124  
 FT VARSELC 268 303  
 SQ SEQUENCE 829 AA; 93697 MW; 7B1E335D4CC809B CRC64;

Query Match 66.0%; Score 33; DB 1; Length 829;  
 Best Local Similarity 85.7%; Pred. No. 29;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 MMHQERK 10  
 DB 488 MMHSERK 494  
 III III

RESULT 7  
 ICEL\_SPOFR STANDARD; PRT; 299 AA.  
 ID ICEL\_SPOFR 89116;  
 AC P89116;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Caspase-1 precursor (EC 3.4.22.-).  
 OS Spodoptera frugiperda (Fall armyworm).  
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 OC Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;  
 OC Diptera; Noctuidae; Noctuidae; Amphipyritinae; Spodoptera.  
 OX NCBI\_TaxID=7108;  
 RN (1)  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE=97153084; PubMed=8999805;  
 RA Ahmad M., Srinivasula S.M., Wang L., Litwack G., Fernandes-Alnemri T.,  
 RA Alnemri E.S.;  
 RT "Spodoptera frugiperda caspase-1, a novel insect death protease that  
 cleaves the nuclear immunophilin FKBP46, is the target of the

```

RT baculovirus antiapoptotic protein p35."
RL J. Biol. Chem. 272:1421-1424(1997).
CC -1- FUNCTION: INVOLVED IN THE ACTIVATION CASCADE OF CASPASES
CC RESPONSIBLE FOR APOPTOSIS EXECUTION (BY SIMILARITY). INHIBITED BY
CC THE BACULOVIRUS ANTI-APOPTOTIC PROTEIN P35. CLEAVES P35 AND
CC NUCLEAR IMMUNOPHILIN FKBP46.
CC -1- SUBUNIT: HETERODIMER OF A 19/18 kDa (P19/P18) AND A 12 kDa (P12)
CC SUBUNIT.
CC -1- PTM: AN AUTOCATALYTIC MECHANISM GENERATES THE TWO ACTIVE SUBUNITS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C14.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U81510; AAC47442.1; -.
DR HSSP: PA2574; 1CP3.
DR MEROPS: C14.015; -.
DR InterPro: IPR002398; ICE.
DR InterPro: IPR002138; ICE_P10.
DR InterPro: IPR001309; ICE_P20.
DR Pfam: PF00655; ICE_P10; 1.
DR Pfam: PF00656; ICE_P20; 1.
DR SMART: SM00115; CASC; 1.
DR PROSITE: PS01122; CASPASE_CYS; 1.
DR PROSITE: PS01121; CASPASE_HIS; 1.
DR PROSITE: PS50207; CASPASE_P10; 1.
DR PROSITE: PS50208; CASPASE_P20; 1.
KM Hydrolase; Thiol protease; Zymogen; Apoptosis.
FT PROPP 1 28
FT CHAIN 29 184 CASPASE-1 SUBUNIT P19/18.
FT PROPP 185 195 POTENTIAL.
FT CHAIN 196 299 CASPASE-1 SUBUNIT P12.
FT ACT_SITE 136 136 BY SIMILARITY.
FT ACT_SITE 178 178 BY SIMILARITY.
SQ SEQUENCE 299 AA; 33527 MW; 99FAFED09B04EDE CRC64;

Query Match 64.0%; Score 32; DB 1; Length 299;
Best Local Similarity 62.5%; Pred. No. 15;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 3 AVAMHQRK 10
Db 272 AMMHQOKO 279

RESULT 8
KSF5_ECOLI STANDARD; PRT; 327 AA.
AC Q47334;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Polysialic acid capsule expression protein kpsf.
GN KSF.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97086613; PubMed=8932302;
RA Simpson D.A., Hammarion T.C., Roberts I.S.;
RT "transcriptional organization and regulation of expression of region 1
RT of the Escherichia coli K5 capsule gene cluster."
RL J. Bacteriol. 178:6466-6474(1996).
CC -1- SIMILARITY: BELONGS TO THE SIS FAMILY. GUT0/KPSF SUBFAMILY.
CC -1- SIMILARITY: CONTAINS 2 CBS DOMAINS.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X95264; CA64561.1; -.
DR InterPro: IPR000644; CBS_domain.
DR InterPro: IPR004800; kpsf.
DR InterPro: IPR001347; SIS.
DR Pfam: PF00571; CBS; 2.
DR Pfam: PF01380; SIS; 1.
DR SMART: SM00116; CBS; 2.
DR TIGRfams: TIGR00393; kpsf; 1.
DR ATP-Binding; Repeat; CBS domain.
KW NP_BIND 63 68 ATP (POTENTIAL).
FT DOMAIN 215 268 CBS 1.
FT DOMAIN 280 327 CBS 2.
SQ SEQUENCE 327 AA; 35563 MW; D0C9A38DFB1E02C4 CRC64;

Query Match 64.0%; Score 32; DB 1; Length 327;
Best Local Similarity 55.6%; Pred. No. 17;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 AVAMHQRK 9
Db 182 AIAIHQRK 190

RESULT 9
CHOR_BACTN STANDARD; PRT; 401 AA.
ID CHOR_BACTN 002550;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Chondro-6-sulfatase regulatory protein.
GN CHOR.
OS Bacteroides thetaiotaomicron.
OC Bacteria; Bacteroidetes; Bacteroides; Bacteroidaceae;
OC Bacteroides.
OX NCBI_TaxID=818;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=46-1, and CS3;
RX MEDLINE=93054330; PubMed=1429442;
RA Cheng Q., Hwa V., Salyers A.A.;
RT "A locus that contributes to colonization of the intestinal tract by
RT Bacteroides thetaiotaomicron contains a single regulatory gene (chor)
RT that links two polysaccharide utilization pathways."
RL J. Bacteriol. 174:7185-7193(1992).
CC -1- FUNCTION: LINKS THE HEPARIN AND THE CHONDROITIN SULFATE
CC UTILIZATION PATHWAYS WHICH CONTRIBUTE TO THE COLONIZATION OF THE
CC INTESTINAL TRACT. MAY HAVE A REGULATOR FUNCTION. REQUIRED FOR
CC THE EXPRESSION OF CHONDRO-6-SULFATASE.
CC -1- SIMILARITY: BELONGS TO THE ASLB/ATSB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L00678; AAA22908.1; -.
DR PIR: A47014; A47014.
KM Transcription regulation; Activator.
SQ SEQUENCE 401 AA; 46432 MW; 43C4BE4018AA3735 CRC64;

```

Query Match 64.0%; Score 32; DB 1; Length 401;  
 Best Local Similarity 66.7%; Pred. No. 21;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 2 VAMMHOERK 10  
 DB 316 VEMMSERO 324

RESULT 10  
 ANA\_DROME STANDARD; PRT; 474 AA.

AC 026307; 09Y524;  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Anachronism protein precursor.  
 GN ANA OR CG8084.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_Taxid=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Eye Imaginal disk;  
 RX MEDLINE=93327423; PubMed=7916657;  
 RA Ebens A.J., Garren H., Chevette B.N.R., Zipursky S.L.;  
 RT "The drosophila anachronism locus: a glycoprotein secreted by glia  
 RT inhibits neuroblast proliferation.";  
 RL Cell 74:15-27(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Berkeley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandal D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brockschtein P., Brotlier P.,  
 RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fostler C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegam C.,  
 RA Jaitli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laso P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Matel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merulov G., Mlishtina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Mount S.M., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,

RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 CC -1- FUNCTION: NEGATIVELY REGULATES PROLIFERATION OF NEURONAL PRECURSOR  
 CC CELLS, THEREBY CONTROLLING THE TIMING OF POSTEMBRYONIC  
 CC NEUROGENESIS.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: SYNTHESIZED IN SOME GLIAL CELLS AND SECRETED.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; S63815; AAB27582.1; -  
 CC EMBL; AC006074; -; NOT\_ANNOTATED\_CDS.  
 CC EMBL; AEO03834; AAF58998.1; -  
 CC FLYBASE; Fgn0011746; ana.  
 CC Glycoprotein; Signal; Developmental protein; Neurogenesis.  
 CC FT SIGNAL 1 33  
 CC FT CHAIN 34 474  
 CC FT DOMAIN 453 462  
 CC FT CARBOHYD 54 54  
 CC FT CARBOHYD 62 62  
 CC FT CARBOHYD 73 73  
 CC FT CARBOHYD 116 116  
 CC FT CARBOHYD 144 144  
 CC FT CARBOHYD 342 342  
 CC FT CONFLICT 388 388  
 CC FT SEQUENCE 474 AA; 54015 MW; FE8A43042B595A1 CR664;  
 CC A -> G (IN REF. 1).  
 CC -----

Query Match 64.0%; Score 32; DB 1; Length 474;  
 Best Local Similarity 55.6%; Pred. No. 26;  
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 2 VAMMHOERK 10  
 DB 216 VSIHDOERK 224

RESULT 11  
 MG1A\_TREPA STANDARD; PRT; 496 AA.

ID MG1A\_TREPA  
 AC 056342;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Galactoside transport ATP-binding protein mg1a homolog.  
 GN MG1A OR TP0685.  
 OS Treponema pallidum.  
 OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.  
 OX NCBI\_Taxid=160;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97080510; PubMed=8921855;  
 RA Porcella S.F., Popova T.G., Hagman K.E., Penn C.W., Radolf J.D.,  
 RA Nordard M.V.;  
 RT A mg1-like operon in Treponema pallidum, the syphilis spirochete.";  
 RL Gene 177:115-121(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.



RC STRAIN-Nichols;  
 RX MEDLINE=98332770; PubMed=9665876;  
 RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,  
 RA Dodson R., Gwin M., Hickey E.K., Clayton R., Ketchum K.A.,  
 RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,  
 RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,  
 RA McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,  
 RA Hatch B., Horst K., Roberts K., Sandusky K., Weldman J., Smith H.O.,  
 RA Venter J.C.;  
 RT "Complete genome sequence of Treponema pallidum, the syphilis  
 RT spirochete.";  
 RL Science 281:375-388(1998).  
 CC - FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM  
 CC FOR GALACTOSIDES. PROBABLY RESPONSIBLE FOR ENERGY COUPLING TO  
 CC THE TRANSPORT SYSTEM (BY SIMILARITY).  
 CC - SUBCELLULAR LOCATION: Membrane-associated (Potential).  
 CC - SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MGLA SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U48416; AAC44585.1; -  
 CC EMBL: AE001242; AAC65648.1; -  
 CC TRNG: TP00685; -  
 DR InterPro: IPR003593; AAA\_ATPase.  
 DR InterPro: IPR003439; ABC\_transport.  
 DR Pfam: PF00005; ABC\_tran; 2.  
 DR ProDom: PD000006; ABC\_transport; 2.  
 DR SMART: SM00382; AAA; 2.  
 DR PROSITE: PS00211; ABC\_TRANSPORTER; 1.  
 KW Transport; Sugar transport; Membrane; ATP-binding; Complete proteome.  
 FT NE\_BIND 37 44 ATP (POTENTIAL).  
 FT SEQUENCE 496 AA; 55191 MW; B6F63D53C5CD11 CRC64;  
 SQ  
 Query Match 64.0%; Score 32; DB 1; Length 496;  
 Best Local Similarity 85.7%; Pred. No. 27;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 VAMH0E 8  
 |||:|  
 Db 81 VAMV0E 87

RESULT 12  
 FRDA\_MOLSU STANDARD; PRT; 656 AA.  
 ID FRDA\_MOLSU  
 AC P17412;  
 DT 01-NOV-1990 (Rel. 15, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Fumarate reductase flavoprotein subunit (EC 1.3.99.1).  
 GN FRDA.  
 OS Wolinella succinogenes.  
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
 OC Wolinella.  
 OX NCBI\_TaxID=844;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91058386; PubMed=2244791;  
 RA Lauterbach F., Koertner C., Albracht S.P., Unden G., Kroeger A.;  
 RT "The fumarate reductase operon of Wolinella succinogenes. Sequence  
 RT and expression of the fda and fda genes.";  
 RL Arch. Microbiol. 154:386-393(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Simon J., Gross R., Ringel M., Schmidt E., Kroeger A.;  
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]

RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS), AND REVISIONS TO 281-289.  
 RX MEDLINE=20052158; PubMed=10586875;  
 RA Lancaster C.R.D., Kroeger A., Auer M., Michel H.;  
 RT "Structure of fumarate reductase from Wolinella succinogenes at 2.2 A  
 RT resolution.";  
 RL Nature 402:377-385(1999).  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (3.1 ANGSTROMS).  
 RX MEDLINE=21145485; PubMed=11248702;  
 RA Lancaster C.R.D., Gross R., Simon J.;  
 RT "A third crystal form of Wolinella succinogenes quinol:fumarate  
 RT reductase reveals domain closure at the site of fumarate reduction.";  
 RL Eur. J. Biochem. 268:1820-1827(2001).  
 CC - CATALYTIC ACTIVITY: Succinate + acceptor = fumarate + reduced  
 CC acceptor.  
 CC - COFACTOR: FAD; COVALENTLY LINKED.  
 CC - SUBUNIT: FUMARATE DEHYDROGENASE FORMS PART OF AN ENZYME COMPLEX  
 CC CONTAINING THREE SUBUNITS: A FLAVOPROTEIN, AN IRON-SULFUR, AND  
 CC CYTOCHROME B-556.  
 CC - SIMILARITY: TO THE FLAVOPROTEIN SUBUNITS OF OTHER SPECIES  
 CC FUMARATE REDUCTASE AND OF SUCCINATE DEHYDROGENASE.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AJ000662; CA04214.1; -  
 CC PIR: S10165; S10165.  
 CC PIR: B44954; B44954.  
 CC PDB: 1QLE; 29-NOV-99.  
 DR InterPro: IPR003953; FAD\_bind2.  
 DR InterPro: IPR003952; FRD/SDH\_FAD.  
 DR InterPro: IPR004112; Succ\_DH\_flav\_C.  
 DR Pfam: PF00890; FAD-binding\_2; 1.  
 DR PROSITE: PS00504; FRD\_SDH\_FAD\_BINDING; 1.  
 KW Oxidoreductase; Electron transport; Flavoprotein; FAD; 3D-structure.  
 FT NE\_BIND 9 23 FAD (AMP PAR) (POTENTIAL).  
 FT BINDING 43 43 FAD (COVALENT).  
 FT ACT\_SITE 257 257 FAD (COVALENT).  
 FT ACT\_SITE 273 273 BY SIMILARITY.  
 FT CONFLICT 281 289 BY SIMILARITY.  
 FT SEQUENCE 656 AA; 72834 MW; DF4D499571412326 CRC64;  
 SQ  
 Query Match 64.0%; Score 32; DB 1; Length 656;  
 Best Local Similarity 50.0%; Pred. No. 37;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 1 AVAMH0ERK 10  
 |::|::|  
 Db 181 AIALIHQDGK 190

RESULT 13  
 PTXA\_HUMAN STANDARD; PRT; 802 AA.  
 ID PTXA\_HUMAN  
 AC P18433; Q14513;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Protein-tyrosine phosphatase alpha precursor (EC 3.1.3.48) (R-PTP-  
 DE alpha).  
 GN PTXA OR PTPA.  
 GN Homo sapiens (human).  
 OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RX MEDLINE-90349565; PubMed2166945;  
 RA Sap J., D'Eustachio P., Givol D., Schlessinger J.;  
 RT "Cloning and expression of a widely expressed receptor tyrosine  
 RT phosphatase.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:6112-6116(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-90384936; PubMed2169617;  
 RA Kaplan R., Morse B., Huebner K., Croce C., Howk R., Ravera M.,  
 RA Ricca G., Jaye M., Schlessinger J.;  
 RT "Cloning of three human tyrosine phosphatases reveals a multigene  
 RT family of receptor-linked protein-tyrosine-phosphatases expressed in  
 RT brain.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:7000-7004(1990).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-91006018; PubMed2170109;  
 RA Krueger N.X., Streuli M., Saito H.;  
 RT "Structural diversity and evolution of human receptor-like protein  
 RT tyrosine phosphatases.";  
 RL EMBO J. 9:3241-3252(1990).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX TISSUE-Kidney;  
 RA MEDLINE-91088320; PubMed2175890;  
 RA Ohagi S., Nishi M., Steiner D.F.;  
 RT "Sequence of a cDNA encoding human LRP (leukocyte common antigen-  
 RT related peptide).";  
 RL Nucleic Acids Res. 18:7159-7159(1990).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-91032191; PubMed2172030;  
 RA Jirik F.R., Janzen N.M., Melhado I.G., Harder K.W.;  
 RT "Cloning and chromosomal assignment of a widely expressed human  
 RT receptor-like protein-tyrosine phosphatase.";  
 RL FEBS Lett. 273:239-242(1990).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-21638749; PubMed11780052;  
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,  
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,  
 RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,  
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,  
 RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,  
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,  
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,  
 RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,  
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,  
 RA Graham D.V., Griffiths C., Griffiths M.N.D., Gilliam R., Hall R.E.,  
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,  
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,  
 RA Leharvalho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,  
 RA Marsh V.L., Martin S.L., McConachie L.J., McLeay K., McMurtry A.A.,  
 RA Milne S.A., Misty D., Moore M.J.F., Mullikin J.C., Nickerson T.,  
 RA Oliver K., Parker A., Patel A., Pearce T.A.V., Peck A.I.,  
 RA Phillips B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,  
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkhen R., Sims S.,  
 RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,  
 RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,  
 RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,  
 RA Wilmink L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
 RA Rogers J.;  
 RL "The DNA sequence and comparative analysis of human chromosome 20.";  
 RL Nature 414:865-871(2001).  
 RT -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O -> protein  
 CC tyrosine + phosphate.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M34668; AAA36528.1; -  
 DR EMBL: X54130; CAA38065.1; -  
 DR EMBL: X54890; CAA38662.1; -  
 DR EMBL: X53364; CAA37447.1; -  
 DR EMBL: AL121905; CAC10337.1; -  
 DR PIR: A36065; A36065.  
 DR PIR: S12049; S12049.  
 DR HSSP: P18052; LYFO.  
 DR GeneW: HGNC:9664; PTPRA.  
 DR MIM: 176884; -  
 DR InterPro: IPR000387; TYR\_phosphatase.  
 DR InterPro: IPR000242; TYR\_PP.  
 DR Pfam: PF00102; Y\_phosphatase; 2.  
 DR PRINTS: PR00700; PRTYPHPHTASE.  
 DR SMART: SM00194; PTPC; 2.  
 DR PROSITE: PS00383; TYR\_PHOSPHATASE\_1; 2.  
 DR PROSITE: PS00056; TYR\_PHOSPHATASE\_2; 2.  
 DR PROSITE: PS00055; TYR\_PHOSPHATASE\_PTP; 2.  
 DR GlycoProtein: Transmembrane; Hydrolase; Phosphorylation; Signal;  
 KW Repeat; Alternative splicing.  
 FT CHAIN 1..19  
 FT SIGNAL 1..19  
 FT DOMAIN 20..142  
 FT TRANSMEM 143..165  
 FT DOMAIN 166..802  
 FT DOMAIN 241..500  
 FT DOMAIN 501..802  
 FT ACT\_SITE 442..442  
 FT ACT\_SITE 732..732  
 FT CARBOHYD 21..21  
 FT CARBOHYD 36..36  
 FT CARBOHYD 68..68  
 FT CARBOHYD 80..80  
 FT CARBOHYD 86..86  
 FT CARBOHYD 104..104  
 FT CARBOHYD 124..124  
 FT CARBOHYD 139..147  
 FT VARSPIC 139..147  
 FT CONFLICT 114..114  
 FT CONFLICT 122..122  
 FT CONFLICT 138..138  
 FT CONFLICT 179..187  
 FT CONFLICT 269..289  
 FT CONFLICT 367..367  
 FT CONFLICT 493..493  
 FT CONFLICT 786..786  
 SQ SEQUENCE 802 AA; 90599 MW; 8B964C3B5B5B32 CRC64;  
 Query Match 64.0%; Score 32; DB 1; Length 802;  
 Best Local Similarity 85.7%; Pred. No. 46;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 OY 4 MMH0ERK 10  
 DB 461 MMH0ERK 467  
 RESULT 14  
 ENTK\_BOVIN STANDARD; PRT: 1035 AA.  
 AC P98072;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Enteropепtidase precursor (EC 3.4.21.9) (Enterokinase).  
 GN PRS57 OR ENTK.  
 OS Bos taurus (Bovine).



```

Query Match      64.0%; Score 32; DB 1; Length 1035;
Best Local Similarity 66.7%; Pred. No. 61;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      2 VAMMHOERK 10
Db      893 IAMHLEMK 901

RESULT 15
HEM_DROME
ID      HEM_DROME      STANDARD;      PRT; 1126 AA.
AC      P53162; Q9YNU8;
DT      01-OCT-1996 (Rel. 34, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Membrane-associated protein Hem (Dhem-2).
GN      HEM OR HEM2 OR CG5837.
OS      Drosophila melanogaster (Fruit fly).
OC      Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC      Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC      Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX      NCBI_TaxID=7227;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Canton-S;
RX      MEDLINE=9531111; PubMed=7643388;
RA      Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA      George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA      Sutton G.G., Mortan J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA      Brandon R.C., Rogers Y.H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA      Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA      Abiri J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA      Balles R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA      Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA      Borkova D., Botchan M.R., Bouck J., Brockstein P., Brothier P.,
RA      Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA      Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA      de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA      Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA      Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA      Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA      Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA      Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA      Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA      Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA      Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA      Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA      Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA      Merkulov G., Mishina N.V., Mody B., Murphy C., Morris J., Moshrefi A.,
RA      Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA      Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA      Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA      Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA      Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA      Spier E., Spirdling A.C., Stapleton M., Strong R., Sun E.,
RA      Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA      Wang Z.-Y., Massarman D.A., Weinstein G.M., Weissbach J.,
RA      Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA      Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA      Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

```

```

RA      Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT      "The genome sequence of Drosophila melanogaster.";
RL      Science 287:2185-2195(2000).
CC      -1- FUNCTION: PLAY A ROLE DURING GROWTH OF THE OOCYTE.
CC      -1- DEVELOPMENTAL STAGE: EXPRESSED MATERNALLY IN THE OOCYTE AND SHOWS
CC      UNIFORM EXPRESSION DURING THE FIRST HALF OF EMBRYOGENESIS, BUT
CC      BECOMES RESTRICTED TO THE BRAIN AND THE NERVOUS SYSTEM DURING LATE
CC      EMBRYOGENESIS.
CC      -1- SIMILARITY: BELONGS TO THE HEM-1/HEM-2 FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X80028; CA56332.1; -.
DR      EMBL; AE003597; AAF51820.1; -.
DR      FlyBase; FBgn0011771; Hem.
KW      Transmembrane.
FT      TRANSMEM 989 1006
SQ      SEQUENCE 1126 AA; 129379 MW; 125FE7177AECC0B5 CNC64;
POTENTIAL.

Query Match      64.0%; Score 32; DB 1; Length 1126;
Best Local Similarity 62.5%; Pred. No. 67;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY      3 AMMHORERK 10
Db      331 ALMHRRER 338

```

Search completed: July 11, 2003, 11:55:21  
 Job time: 1.95634 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:44:31 ; Search time 4.3659 Seconds

(without alignments)  
471.946 Million cell updates/sec

Title: US-10-091-442-9

Sequence: 1 AVAMMHQERK 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: SPTREMBL\_21:\*

1: sp.archaea:\*

2: sp.bacteria:\*

3: sp.fungi:\*

4: sp.human:\*

5: sp.invertebrate:\*

6: sp.mammal:\*

7: sp.mhc:\*

8: sp.organelle:\*

9: sp.phage:\*

10: sp.plant:\*

11: sp rodent:\*

12: sp.virus:\*

13: sp.vertebrate:\*

14: sp.unclassified:\*

15: sp.virus:\*

16: sp.bacteriap:\*

17: sp.archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	82.0	380	11	Q920J5
2	38	76.0	245	5	Q9NB28
3	37	74.0	380	11	Q9D695
4	36	72.0	553	2	Q93AG2
5	35	70.0	212	13	O42614
6	35	70.0	376	5	P91376
7	35	70.0	500	13	Q8UV19
8	35	70.0	578	10	O23030
9	35	70.0	602	10	O22783
10	35	70.0	1561	11	O88286
11	35	70.0	1568	4	O95785
12	34	68.0	82	4	Q91509
13	34	68.0	82	11	O9CR20
14	34	68.0	695	10	O8W2R3
15	33	66.0	200	10	O9FUS9
16	33	66.0	275	16	O50755

17	33	66.0	281	16	Q8YVK5	Q8YVK5 anabaena sp
18	33	66.0	303	2	Q9RC27	Q9RC27 bacillus sp
19	33	66.0	318	10	Q94JB7	Q94JB7 oryza sativ
20	33	66.0	349	5	Q9USN7	Q9USN7 leishmania
21	33	66.0	366	16	Q985Z3	Q985Z3 rhizobium l
22	33	66.0	395	13	Q90231	Q90231 alligator m
23	33	66.0	536	4	O8TE48	O8TE48 homo sapien
24	33	66.0	642	4	O96K06	O96K06 homo sapien
25	33	66.0	761	5	O95Y05	O95Y05 leishmania
26	33	66.0	793	11	O91V35	O91V35 mus muscula
27	33	66.0	1018	5	Q95TW8	Q95TW8 drosophila
28	33	66.0	1232	5	Q9VFI9	Q9VFI9 drosophila
29	32	64.0	161	10	O8VWR4	O8VWR4 arabidopsis
30	32	64.0	229	10	O9FZJ8	O9FZJ8 arabidopsis
31	32	64.0	252	16	O8YGA5	O8YGA5 brucella me
32	32	64.0	298	16	P71702	P71702 mycobacteri
33	32	64.0	371	13	Q90ZL9	Q90ZL9 xenopus lae
34	32	64.0	473	13	Q91B97	Q91B97 potamoctrygo
35	32	64.0	486	16	O91278	O91278 pseudomonas
36	32	64.0	496	2	O56341	O56341 treponema p
37	32	64.0	497	2	O9ZGA2	O9ZGA2 treponema d
38	32	64.0	526	5	O8T823	O8T823 drosophila
39	32	64.0	581	5	Q9VNM4	Q9VNM4 drosophila
40	32	64.0	708	4	O96S08	O96S08 homo sapien
41	32	64.0	738	10	O9XEM3	O9XEM3 oryza sativ
42	32	64.0	802	4	O96TDP	O96TDP homo sapien
43	32	64.0	807	13	O91969	O91969 gallus gall
44	32	64.0	1220	4	O9P252	O9P252 homo sapien
45	32	64.0	1972	4	Q9UIF8	Q9UIF8 homo sapien

## ALIGNMENTS

RESULT 1	ID	Q920J5	PRELIMINARY:	PRT:	380 AA.
AC	Q920J5	01-DEC-2001 (TREMURel. 19, Created)			
DT	01-DEC-2001 (TREMURel. 19, Last sequence update)				
DT	01-MAR-2002 (TREMURel. 20, Last annotation update)				
DE	Meglin.				
OS	Rattus norvegicus (Rat).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
OX	NCBI_TaxID=10116;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE-2136806; PubMed-11473647;				
RA	Nagaku M., Miyata T., Suzuki D., Umezono T., Hashimoto T., Wada T.,				
RA	Yagi M., Nagano N., Inagi R., Kurokawa K.;				
RT	"Cloning of rodent meglin revealed its up-regulation in				
RT	mesangiolipofoliferative nephritis.";				
RL	Kidney Int. 60:641-652(2001).				
CC	-1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.				
DR	EMBL: AF105329; AAL16769.1; -				
DR	InterPro: IPR00215; Serpin.				
DR	Pfam: PF00079; serpin.1.				
DR	PROSITE: PS00284; SERPIN; UNKNOWN_1.				
KW	Serpin.				
SO	SEQUENCE 380 AA; 42821 MW; D8076CABEE2C2FBC CRC64;				
QY	Query Match	82.0%;	Score 41;	DB 11;	Length 380;
DB	Best Local Similarity	80.0%;	Pred. No. 1.9;		
	Matches 8;	Conservative 1;	Mismatches 1;	Indels 0;	Gaps 0;
	1 AVAMMHQERK 10				
	204 AVAMMHQERK 213				
RESULT 2	Q9NB28				

ID Q9NB28 PRELIMINARY; PRT; 245 AA.  
 AC Q9NB28;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE HMG box transcription factor Amphiox/2/3.  
 DE Branchiostoma floridae (Florida lancelet) (Amphioxus);  
 OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;  
 OC Branchiostoma.  
 OX NCBI\_TaxID=7739;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Holland L.Z., Schubert M., Holland N.D., Neuman T.;  
 RT "Evolutionary conservation of the presumptive neural plate markers  
 RT Amphiox/2/3 and AmphibNeurogenin in the invertebrate chordate  
 RT amphioxus.";  
 RL Dev. Biol. 0:0-0(2000).  
 DR EMBL: AF271787; AAF81765.1; -.  
 DR HSSP: Q05066; 1HRX.  
 DR InterPro: IPR000135; Highmobily\_12.  
 DR InterPro: IPR000910; HMG\_12\_box.  
 DR Pfam: PF00505; HMG\_box.1.  
 DR PRINTS: PR00886; HIGHMOBLTY12.  
 DR SMART: SM00398; HMG.1.  
 SQ SEQUENCE 245 AA; 28158 MW; F4B58A2P4C162299 CRC64;

Query Match 76.0%; Score 38; DB 5; Length 245;  
 Best Local Similarity 100.0%; Pred. No. 5.2;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 VMMHOERK 10  
 11111111  
 DB 6 VMMHOERK 12

RESULT 3  
 Q9D695 PRELIMINARY; PRT; 380 AA.  
 AC Q9D695;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE 4631416M05Rik protein (Megasin).  
 GN SERPINB7 OR 4631416M05Rik.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=SKIN;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Felschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boilelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasakih H., Sato K., Schenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyo-Oka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wyszynski A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=21368006; PubMed=11473647;  
 RA Nagaku M., Miyata T., Suzuki D., Umezono T., Hashimoto T., Wada T.,  
 RA Yagi M., Nagano N., Inagi R., Kurokawa K.;  
 RT "Cloning of rodent mesgin revealed its up-regulation in  
 RT mesangio proliferative nephritis.";  
 RL Kidney Int. 60:641-652(2001).  
 DR EMBL: AK014524; BAB29410.1; -.  
 DR EMBL: AF105328; AAL16768.1; -.  
 DR HSSP: P05121; 1ATC.  
 DR MGD: MGI:2151053; Serpinb7.  
 DR InterPro: IPR000215; Serpin.  
 DR Pfam: PF00079; serpin.1.  
 DR SMART: SM00093; SERPIN.1.  
 DR PROSITE: PS00284; SERPIN; UNKNOWN.1.  
 SQ SEQUENCE 380 AA; 43050 MW; C9240272BCFB9CFA CRC64;

Query Match 74.0%; Score 37; DB 11; Length 380;  
 Best Local Similarity 77.8%; Pred. No. 13;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 VMMHOERK 10  
 11111111  
 DB 205 VMMHOERK 213

RESULT 4  
 Q93AG2 PRELIMINARY; PRT; 553 AA.  
 AC Q93AG2;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE FAD-monoxygenase.  
 GN CADA.  
 OS Pseudomonas putida.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=303;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FO07;  
 RA Trevino-Quintanilla L.G., Soberon-Chavez G.;  
 RT "Isolation and characterization of a Pseudomonas putida strain able to  
 RT degrade chloranilic acid (2,5-dichloro-3,6-dihydroxy benzo-1,4-  
 RT quione).";  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF421356; AAL16082.1; -.  
 DR InterPro: IPR000733; Flay\_monoxygenase.  
 DR InterPro: IPR002938; MoxY\_FAD\_binding.  
 DR InterPro: IPR000205; NAD\_binding.  
 DR InterPro: IPR000594; Thlf domain.  
 DR Pfam: PF01494; FAD\_binding\_3; 1.  
 DR Pfam: PF01360; Monoxygenase; 1.  
 KW Monoxygenase.  
 SQ SEQUENCE 553 AA; 61430 MW; 54BB707D639E0C6E CRC64;

Query Match 72.0%; Score 36; DB 2; Length 553;  
 Best Local Similarity 77.8%; Pred. No. 31;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AVAMHOER 9  
 11111111  
 DB 369 AVAMHOER 377

RESULT 5  
 Q42614 PRELIMINARY; PRT; 212 AA.  
 AC Q42614;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE Antithrombin III (Fragment).  
 OS Cyprinus carpio (Common carp).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Cyprinus.  
 OX NCBI\_Taxid=7962;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Gracey A.Y., Colonna-Romano S., Tikku P.E., Cossins A.R., Maresca B.;  
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.  
 DR EMBL: AF006495; AAB61755.1; -;  
 DR EMBL: AF006494; AAC19409.1; -;  
 DR HSSP: P01008; IBR8.  
 DR InterPro: IPR000215; Serpin.  
 DR Pfam: PF00079; serpin.1.  
 DR SMART: SM00093; SERPIN.1.  
 DR PROSITE: PS00284; SERPIN; UNKNOWN\_1.  
 KM Serpin.  
 FT NON\_TER  
 SQ SEQUENCE 212 AA; 23875 MW; 7F185103CF8BAFF0 CRC64;  
 Query Match 70.0%; Score 35; DB 13; Length 212;  
 Best Local Similarity 77.8%; Pred. No. 19;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 OY 2 VAMMHQERK 10  
 Db 27 VPMVQERK 35  
 ID P91376 PRELIMINARY; PRT; 376 AA.  
 AC P91376;  
 DT 01-MAY-1997 (TREMBLrel. 03, Created)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Hypothetical 42.4 kDa protein.  
 GN K11H12.9  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Pelodermidae; Caenorhabditis.  
 OX NCBI\_Taxid=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA None;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium.";  
 RL Science 282:2012-2018(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Bradshaw H.;  
 RT "The sequence of C. elegans cosmid K11H12.";  
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Materon R.;  
 RT "Direct Submision.";  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U88168; AAC24403.2; -;  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR002290; Ser\_thr\_pkinase.  
 DR Pfam: PF00069; pkinase.1.  
 DR PRODOM: PD000001; Euk\_pkinase.2.  
 DR SMART: SM00220; S\_TKc.1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM.1.  
 DR Hypothetical protein.  
 KW SEQUENCE 376 AA; 42391 MW; 60F460A4685000F8 CRC64;

Query Match 70.0%; Score 35; DB 5; Length 376;  
 Best Local Similarity 100.0%; Pred. No. 34;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 AVAMMHQ 7  
 Db 208 AVAMMHQ 214  
 ID O8UV19 PRELIMINARY; PRT; 500 AA.  
 AC O8UV19;  
 DT 01-MAR-2002 (TREMBLrel. 20, Created)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Hypothetical 63.6 kDa protein (Fragment).  
 GN C16ORF35.  
 OS Sphaerooides nephelus.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
 OC Tetraodontidae; Sphaerooides.  
 OX NCBI\_Taxid=39110;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21096913; PubMed=11157800;  
 RA Flint J., Tufarelli C., Peden J., Clark K., Daniels R.J., Hardison R.,  
 RA Miller W., Phillipsen S., Tan-Un K.C., McMorro T., Frampton J.,  
 RA Alter B.P., Frischauf A.M., Higgs D.R.;  
 RT "Comparative genome analysis delimits a chromosomal domain and  
 RT identifies key regulatory elements in the alpha globin cluster.";  
 RL Hum. Mol. Genet. 10:371-382(2001).  
 DR EMBL: AY016023; AAL73050.1; -;  
 DR InterPro: IPR005365; UPF0171.  
 DR Pfam: PF03666; UPF0171.1.  
 FT NON\_TER  
 FT NON\_TER  
 SQ SEQUENCE 500 AA; 56492 MW; 8545420A768B1879 CRC64;  
 Query Match 70.0%; Score 35; DB 13; Length 500;  
 Best Local Similarity 50.0%; Pred. No. 45;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
 OY 1 AVAMMHQERK 10  
 Db 89 AIALQHEERR 98  
 ID O23030 PRELIMINARY; PRT; 578 AA.  
 AC O23030;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE TIG11.21 protein.  
 GN TIG11.21.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eufosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_Taxid=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV, COLUMBIA;  
 RA Osborne B.I., Vysotskaya V.S., Toriumi M., Yu G., Oji O., Liu S.,  
 RA Li J., Hoang L., Buehler E., Conway A.B., Conway A.R., Dewar K.,  
 RA Feng J., Kim C., Kurtz D., Li Y., Shinn P., Sun H., Davis R.W.,  
 RA Ecker J.R., Federspiel N.A., Theologis A.;  
 RT "The sequence of BAC TIG11 from Arabidopsis thaliana chromosome 1.";  
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.

```

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA Theologis A.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC002376; AAB0639.1; -.
DR InterPro: IPR000008; C2.
DR Pfam: PF00168; C2; 1.
DR SMART: SM00239; C2; 1.
SQ SEQUENCE 578 AA; 64860 MW; 5C3D8369A580C9D2 CRC64;

Query Match 70.0%; Score 35; DB 10; Length 578;
Best Local Similarity 70.0%; Pred. No. 52;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 AVAMHQRK 10
   111:1111
DB 29 AVAMVHSRK 38

RESULT 9
ID 022783 PRELIMINARY; PRT; 602 AA.
AC 022783;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Atg933320 protein.
GN Atg933320.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA "MEDLINE-20083487; PubMed-10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Unayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RA "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RA thaliana.";
RT Nature 402:761-768(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA Lin X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC002332; AAB0650.1; -.
DR InterPro: IPR000008; C2.
DR Pfam: PF00168; C2; 1.
DR SMART: SM00239; C2; 1.
SQ SEQUENCE 602 AA; 66580 MW; 8045E5A2B895C26A CRC64;

Query Match 70.0%; Score 35; DB 10; Length 602;
Best Local Similarity 70.0%; Pred. No. 55;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 AVAMHQRK 10
   111:1111
DB 31 AVAMVHSRK 40

RESULT 10
ID 088286 PRELIMINARY; PRT; 1561 AA.
AC 088286;
DT 01-NOV-1998 (TREMBLrel. 08, Created)

```

```

DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE WIZL.
GN WIZ OR WIZ.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RA Matsumoto K., Ishii N., Yoshida S., Shiosaka S., Wanaka A.,
RA Tohyama M.;
RT "Molecular Cloning and Distinct Developmental Expression Pattern of
RT Sliced Forms of A Novel Zinc Finger Gene wiz in the Cerebellum.";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL: AB012265; BAA32790.1; -.
DR MGD: MG1:1332638; WIZ.
DR InterPro: IPR000822; znf_C2H2.
DR Pfam: PF00096; znf_C2H2; 10.
DR PRINTS: PR00048; ZINCINGER.
DR SMART: SM00355; znf_C2H2; 10.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 8.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 7.
KW DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.
SQ SEQUENCE 1561 AA; 171243 MW; 8A010A7B5374CF6 CRC64;

Query Match 70.0%; Score 35; DB 11; Length 1561;
Best Local Similarity 60.0%; Pred. No. 14e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 AVAMHQRK 10
   1:1:1111
DB 604 AMALKHEERK 613

RESULT 11
ID 095785 PRELIMINARY; PRT; 1568 AA.
AC 095785;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Human homolog of MUS MUSCULUS WIZL protein (Human homolog of MUS
DE MUSCULUS WIZ protein) (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA lamerdin J.E., McCreedy P.M., Skowronski E., Viswanathan V.,
RA Burkhardt-Schultz K., Gordon L., Dias J., Ramirez M., Stillwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Gaines J.,
RA Dangnan L., Erlar A., Christensen M., Amico-Keller G., Coefield J.,
RA Attix C., Andreise T., Frankheim M., Amico-Keller G., Coefield J.,
RA Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Krommiller B.,
RA Arellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A.,
RA Olsen A.S., Carrano A.V.;
RT "Sequence analysis of a 3.5 Mb region in 19p13.1 between OLF1 and
RT JUND.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 783-1568 FROM N.A.
RA lamerdin J.E., McCreedy P.M., Skowronski E., Viswanathan V.,
RA Burkhardt-Schultz K., Gordon L., Dias J., Brower A., Ramirez M.,
RA Stillwagen S., Phan H., Velasco N., Do L., Regala W., Terry A.,
RA Gaines J., Dangnan L., Erlar A., Christensen M., Georgescu A.,
RA Avila J., Liu S., Attix C., Andreise T., Frankheim M.,
RA Amico-Keller G., Coefield J., Duarte S., Lucas S., Bruce R.,
RA Thomas P., Quan G., Krommiller B., Arellano A., Sanders C., Ow D.,
RA Nolan M., Trong S., Kobayashi A., Olsen A.S., Carrano A.V.;

```



RT "Sequence analysis of an ~1.5 Mb contig in 19p13.1 between OLFR and  
 RT D19S885." ;  
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC007059; AAD19818.1; -  
 DR EMBL: AC006128; AAC97985.1; -  
 DR EMBL: AC007059; AAD19817.1; -  
 DR InterPro: IPR000822; 2nf.C2H2.  
 DR Pfam: PF00096; zf.C2H2; 10.  
 DR SMART: SMO0355; 2nf.C2H2; 10.  
 DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 8.  
 DR PROSITE: PS50157; ZINC\_FINGER\_C2H2\_2; 7.  
 KW DNA-binding; Metal-binding; Zinc-finger.  
 FT NON\_TER  
 SQ SEQUENCE 1568 AA; 170077 MW; 7AF0D34D45F565D0 CRC64;

OY 1 AVAMHOERK 10  
 Db 610 AMALKEERK 619

Query Match 70.0%; Score 35; DB 4; Length 1568;  
 Best Local Similarity 60.0%; Pred. No. 1,4e+02;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

RESULT 12  
 ID 09Y5U9 PRELIMINARY; PRT; 82 AA.  
 AC 09Y5U9;  
 DT 01-NOV-1999 (TReMBLrel. 12, Created)  
 DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)  
 DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)  
 DE HSPC039 protein (Hypothetical 9.0 kDa protein) (Immediate early  
 response 3 interacting protein) (Similar to HSPC039 protein).  
 CN DRZFP64B1471.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ye M., Zhang Q., Zhou J., Shen Y., Guan Z., Wu X., Fan H., Mao H.,  
 RA Dai M., Huang Q., Chen S., Chen Z.;  
 RT "Human HSPC039 mRNA, complete cds." ;  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-ADRENAL GLAND;  
 RA Huang C., Zhang C., Wu T., Peng Y., Gu Y., Gu W., Jiang C., Li Y.,  
 RA Han Z., Wang Y., Chen Z., Fu G.;  
 RT "A novel gene expressed in human adrenal gland." ;  
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN;  
 RA Bloecher H., Boecher M., Brandt P., Mewes H.W., Gassenhuber J.,  
 RA Wiemann S.;  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LIVER;  
 RA Yiu W.H., Kok L.D.S., Fung K.P., Lee C.Y., Tsui S.K.W., Waye M.M.Y.;  
 RT "Characterization of immediate early response 3 interacting protein  
 (IERIP)." ;  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN;  
 RA Strausberg R.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-MYELOID LEUKEMIA CELLS;  
 RA Strausberg R.;

RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF125100; AAD39917.1; -  
 DR EMBL: AF164798; AAF80762.1; -  
 DR EMBL: AL136667; CAB66602.1; -  
 DR EMBL: AF371963; AAK53816.1; -  
 DR EMBL: BC010888; AAH10888.1; -  
 DR EMBL: BC017391; AAH17391.1; -  
 KW Transmembrane.  
 SQ SEQUENCE 82 AA; 8969 MW; 2015B211F0AF580 CRC64;

OY 1 AVAMHOER 9  
 Db 18 AIALVHEER 26

Query Match 68.0%; Score 34; DB 4; Length 82;  
 Best Local Similarity 55.6%; Pred. No. 12;  
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

RESULT 13  
 ID 09CR20 PRELIMINARY; PRT; 82 AA.  
 AC 09CR20;  
 DT 01-JUN-2001 (TReMBLrel. 17, Created)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)  
 DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)  
 DE 1110057H19R1k protein (RIKEN cDNA 1110057H19 gene).  
 CN 1110057H19R1K.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=TONGUE, EMBRYO, AND PANCREAS;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Aikawa K., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., But C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nodone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Winking L.,  
 RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawai J., Kontuski S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection." ;  
 RL Nature 409:685-690(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Strausberg R.;  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AK009116; BAB26081.1; -  
 DR EMBL: AK004281; BAB23248.1; -  
 DR EMBL: AK007388; BAB25005.1; -  
 DR EMBL: AK007862; BAB25314.1; -  
 DR EMBL: BC021512; AAH21512.1; -  
 DR MGI:1913441; 1110057H19R1K.  
 SQ SEQUENCE 82 AA; 9017 MW; 3015F654F1EBB080 CRC64;

OY 1 AVAMHOER 9

Query Match 68.0%; Score 34; DB 11; Length 82;  
 Best Local Similarity 55.6%; Pred. No. 12;  
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 18 AVALHEER 26

## RESULT 14

Q8W2R3 PRELIMINARY; PRT; 695 AA.  
 AC Q8W2R3;  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Putative wall-associated protein kinase.  
 GN OSUBA0028C16.9.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Erihartoideae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Wang R.A., Yu Y., Soderlund C., Chen M., Kim H.-R., Rambo T.,  
 RT Saski C., Henry D., Oates R., Simmons J.;  
 "Rice Genomic Sequence."  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC098565; AAL69427.1; -  
 DR InterPro: IPR000152; Asx\_hydroxyl.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR001881; EGF\_Ca.  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR002290; Ser\_thr\_pkinase.  
 DR InterPro: IPR001245; Tyr\_pkinase.  
 DR Pfam: PF00069; pkinase; 1.  
 DR ProDom: PD000001; Euk\_pkinase; 1.  
 DR SMART: SM00220; S\_TKc; 1.  
 DR SMART: SM00219; TyKc; 1.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; UNKNOWN\_1.  
 DR PROSITE: PS01187; EGF\_CA; UNKNOWN\_1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; UNKNOWN\_1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; UNKNOWN\_1.  
 KW kinase.  
 SQ SEQUENCE 695 AA; 77048 MW; D9DFE3A74A655A5 CRC64;  
 Query Match 68.0%; Score 34; DB 10; Length 695;  
 Best Local Similarity 85.7%; Pred. No. 1e+02; 0; Mismatches 0; Indels 0; Gaps 0;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 4 MMHQERK 10  
 Db 382 MMHQERK 388  
 RESULT 15  
 Q9FU59 PRELIMINARY; PRT; 200 AA.  
 AC Q9FU59;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE P0013F10.15 protein.  
 GN P0013F10.15.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Erihartoideae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC  
 clone:P0013F10.";

RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AP002523; BAB17069.1; -  
 SQ SEQUENCE 200 AA; 22103 MW; 01D469471C059FD8 CRC64;

Query Match 66.0%; Score 33; DB 10; Length 200;  
 Best Local Similarity 60.0%; Pred. No. 48;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 AVAMMHQERK 10  
 Db 138 AVAMMHQERK 147

Search completed: July 11, 2003, 11:59:10  
 Job time : 7.3659 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:42:16 ; Search time 218.046 Seconds  
(Without alignments) 232.223 Million cell updates/sec

Title: US-10-091-442-34

Perfect score: 1959

Sequence: 1 MASLAANAAEFCEFLFREMD.....FLFYRKDDIILFGKVSCLP 380

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: A\_Geneseq\_101002.\*

2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*

3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*

4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*

5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*

6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*

7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*

8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*

9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*

10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*

11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*

12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*

13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*

14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*

15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*

16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*

17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*

18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*

19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*

20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*

21: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*

22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*

23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*

24: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result NO.	Score	Query Match	Length	ID	Description
1	1959	100.0	380	AA8379	Human megakaryocyte
2	1959	100.0	380	AA108254	Human megaslin prote
3	1959	100.0	380	AA824142	Human megaslin prote
4	1959	100.0	380	AA83075	Human megaslin prote
5	1488	76.0	380	AA64286	Rat megaslin protein
6	1477	75.4	380	AA108255	Rat megaslin protein
7	1477	75.4	380	AA824150	Rat megaslin protein
8	1477	75.4	380	AA83076	Rat megaslin protein
9	1458	74.4	368	AA108256	Mouse megaslin prote
10	1458	74.4	368	AA824151	Mouse megaslin prote

11	1458	74.4	368	AA83077	Murine megaslin prot
12	796.5	40.7	379	AA15155	Human prostate ser
13	771.5	39.4	395	AA47207	Human NOVI protein
14	769	39.3	394	AB53280	Human polypeptide
15	752	38.4	390	AA825276	SCC antigen. Synt
16	752	38.4	390	AA132077	Hepatitis B virus
17	752	38.4	390	AA125928	Human SCCA1 protei
18	752	38.4	390	AA125928	Human squamous cel
19	750	38.3	390	AA132078	Hepatitis B virus
20	748	38.2	390	AA125927	Human SCCA2 protei
21	748	38.2	390	AA125927	Human squamous cel
22	744	38.0	390	AA152421	Psoriasis-like type
23	740	37.8	390	AA152421	Human prostate ser
24	738.5	37.7	617	AA15156	Human elastase inh
25	722.5	36.9	437	AA133755	Human cancer assoc
26	722.5	36.9	437	AA133755	Human cancer assoc
27	717.5	36.6	379	AA104159	Human lung cancer-
28	715.5	36.5	391	AA141029	Human lung cancer-
29	715.5	36.5	391	AA11314	Human headpin (for
30	715.5	36.5	391	AA12651	Human lung tumor
31	715.5	36.5	391	AA141030	Human lung tumor a
32	715	36.5	400	AA11315	Human lung cancer-
33	715	36.5	400	AA11315	Human lung tumor
34	715	36.5	400	AA125926	Human hurpin prote
35	714.5	36.5	391	AA125926	Human serine prote
36	714.5	36.5	391	AA10954	Human headpin (hea
37	714.5	36.5	391	AA10954	Human headpin (hea
38	702	35.8	400	AA12652	Novel human diagno
39	690	35.2	617	AA12652	Plasminogen activa
40	673	34.4	382	AA12144	Human novel secret
41	672.5	34.3	340	AA16928	Plasminogen activa
42	671	34.3	392	AA12143	Sequence of human
43	664.5	33.9	415	AA10473	Encodes Plasminoge
44	664.5	33.9	415	AA10473	
45	664.5	33.9	415	AA10921	

## ALIGNMENTS

RESULT 1	AA848379	standard; Protein: 380 AA.
ID	AA848379	
XX	AA848379;	
AC	16-AUG-1994 (first entry)	
DT		
XX		
XX		
DE	Human megakaryocyte differentiation factor.	
XX		
XX	Human megakaryocyte differentiation factor; MGF; thrombopoietin;	
KN	haematopoietic stimulating factor; thrombocytopoietin; platelet;	
KN	bone marrow transplantation; cancer chemotherapy.	
XX		
OS	Homo sapiens.	
XX		
PN	EP583884-A.	
XX		
PD	23-FEB-1994.	
XX		
PF	19-JUL-1993; 93EP-030564.	
XX		
PR	17-JUL-1992; 92JP-0212305.	
PR	04-MAR-1993; 93JP-0067339.	
XX		
PA	(SUNR ) SUNTORY LTD.	
PA	(TSUJ/) TSUJIMOTO M.	
PI	Ishida N, Iwasa F, Kurihara T, Miura K, Nakazato H;	
PI	Tsujimoto M, Tsuruoka N, Yamaguchi N, Yamachi K;	
XX		
DR	WPI: 1994-058782/08.	
DR	N-PSDB: AA056670.	

XX New megakaryocyte differentiation factor - isolated from human  
 PT epidermoid carcinoma cells, used to treat conditions involving a  
 PT decrease in platelets  
 XX  
 PS Claim 7; Page 30-32; 47pp; English.  
 XX  
 CC Human MDF can be isolated from a culture of human epidermoid  
 CC carcinoma A431 cells in protein-free medium. The MDF stimulates  
 CC differentiation of megakaryocytes from myeloid cells in the presence  
 CC of IL-3. The MDF acts in vivo as a thrombopoietin making it useful  
 CC for treatment of diseases involving a decrease in platelet number  
 CC (esp. thrombocytopenia) such as occurs in bone marrow  
 CC transplantation and in chemotherapy. MDF has mol.wt. 55-57kD by  
 CC SDS-PAGE and exhibits an isoelectric point of 6.0-7.0.  
 CC  
 XX  
 SQ Sequence 380 AA;  
 Query Match 100.0%; Score 1959; DB 15; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 3.6e-169;  
 Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MASLAANAEEFCNLFREMDNNGNGNVFFSSLSLFAALALVRLGAQDDSLSDIKLHV 60  
 DB 1 MASLAANAEEFCNLFREMDNNGNGNVFFSSLSLFAALALVRLGAQDDSLSDIKLHV 60  
 QY 61 NTASGYGNSSNSOSGLOSLKRVFSDINASHKDYDLSVNGLFAEKYGFHKDYIECAEK 120  
 DB 61 NTASGYGNSSNSOSGLOSLKRVFSDINASHKDYDLSVNGLFAEKYGFHKDYIECAEK 120  
 QY 121 LYAKKERVDFTHLEDTNRNINKWVENETHGKIKNVIGGGISSAAMVLYNAVYFKGK 180  
 DB 121 LYAKKERVDFTHLEDTNRNINKWVENETHGKIKNVIGGGISSAAMVLYNAVYFKGK 180  
 QY 181 WQSAFTKSETINCHFKSPKSGKAVAMHQRKFNLSVIDPMSKLTLELYNGGINNYVL 240  
 DB 181 WQSAFTKSETINCHFKSPKSGKAVAMHQRKFNLSVIDPMSKLTLELYNGGINNYVL 240  
 QY 241 LPENDLSEIENKLTFOMLMWTNPRRTSKYVEVEFFQPKIEKNYKQYLRLGLKDI 300  
 DB 241 LPENDLSEIENKLTFOMLMWTNPRRTSKYVEVEFFQPKIEKNYKQYLRLGLKDI 300  
 QY 301 DESKADLSGASGRLYISRMHKSYLEVEEGTEATAGSNIVEKOLPOSTLFRADHP 360  
 DB 301 DESKADLSGASGRLYISRMHKSYLEVEEGTEATAGSNIVEKOLPOSTLFRADHP 360  
 QY 361 FLFVIRKDDITLFSGKVS 380  
 DB 361 FLFVIRKDDITLFSGKVS 380  
 RESULT 2  
 ID AAY08254 standard; Protein: 380 AA.  
 AC AAY08254;  
 DT 14-JUL-1999 (first entry)  
 DE Human megasin protein.  
 KW Megsin; mesangial cell; treatment; diagnosis; disease; Iga nephropathy;  
 KW human; rat; murine.  
 OS Homo sapiens.  
 PN W09915652-A1.  
 PD 01-APR-1999.  
 PF 22-SEP-1998; 98WO-JP04269.  
 PR 22-SEP-1997; 97JP-0275302.

XX (KURO/) KUROKAWA K.  
 PA (MIYA/) MIYATA T.  
 XX  
 PI Miyata T;  
 XX  
 DR WPI; 1999-276983/23.  
 DR N-PSDB; AAX56712.  
 PS Megsin protein expressed specifically in mesangial cells  
 XX  
 CC Claim 1; Page 62-64; 100pp; Japanese.  
 CC This invention describes the isolation of novel megasin nucleic acid and  
 CC proteins from human, rat and mouse tissue. This protein is expressed  
 CC specifically in mesangial cells. The products of the invention are  
 CC useful for the treatment and diagnosis of diseases involving mesangial  
 CC cells, such as Iga nephropathy.  
 CC  
 XX  
 SQ Sequence 380 AA;  
 Query Match 100.0%; Score 1959; DB 20; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 3.6e-169;  
 Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MASLAANAEEFCNLFREMDNNGNGNVFFSSLSLFAALALVRLGAQDDSLSDIKLHV 60  
 DB 1 MASLAANAEEFCNLFREMDNNGNGNVFFSSLSLFAALALVRLGAQDDSLSDIKLHV 60  
 QY 61 NTASGYGNSSNSOSGLOSLKRVFSDINASHKDYDLSVNGLFAEKYGFHKDYIECAEK 120  
 DB 61 NTASGYGNSSNSOSGLOSLKRVFSDINASHKDYDLSVNGLFAEKYGFHKDYIECAEK 120  
 QY 121 LYAKKERVDFTHLEDTNRNINKWVENETHGKIKNVIGGGISSAAMVLYNAVYFKGK 180  
 DB 121 LYAKKERVDFTHLEDTNRNINKWVENETHGKIKNVIGGGISSAAMVLYNAVYFKGK 180  
 QY 181 WQSAFTKSETINCHFKSPKSGKAVAMHQRKFNLSVIDPMSKLTLELYNGGINNYVL 240  
 DB 181 WQSAFTKSETINCHFKSPKSGKAVAMHQRKFNLSVIDPMSKLTLELYNGGINNYVL 240  
 QY 241 LPENDLSEIENKLTFOMLMWTNPRRTSKYVEVEFFQPKIEKNYKQYLRLGLKDI 300  
 DB 241 LPENDLSEIENKLTFOMLMWTNPRRTSKYVEVEFFQPKIEKNYKQYLRLGLKDI 300  
 QY 301 DESKADLSGASGRLYISRMHKSYLEVEEGTEATAGSNIVEKOLPOSTLFRADHP 360  
 DB 301 DESKADLSGASGRLYISRMHKSYLEVEEGTEATAGSNIVEKOLPOSTLFRADHP 360  
 QY 361 FLFVIRKDDITLFSGKVS 380  
 DB 361 FLFVIRKDDITLFSGKVS 380  
 RESULT 3  
 ID AAB24142 standard; Protein: 380 AA.  
 AC AAB24142;  
 DT 30-JAN-2001 (first entry)  
 DE Human megasin protein sequence SEQ ID NO:2.  
 KW Megsin; mesangium-predominant gene; seipin regulated; nephropathy;  
 KW Iga; immunoglobulin A; detection; renal function; renal disorder;  
 KW diagnosis; biological sample; blood; urine.  
 OS Homo sapiens.  
 PN W0200057189-A1.  
 PD 28-SEP-2000.

XX 17-MAR-2000; 2000WO-JP01646.  
 XX  
 XX 19-MAR-1999; 99JP-0075305.  
 PR 28-OCT-1999; 99JP-0306623.  
 XX  
 PA (KURO/) KUROKAWA K.  
 PA (FUSO ) FUSO PHARM IND LTD.  
 PA (MIYA/) MIYATA T.  
 PI Miyata T.  
 XX  
 XX WPI: 2000-611642/58.  
 DR N-PSDB: AAA99294.  
 XX  
 PT Evaluating renal function comprises assaying mesgin protein in  
 PT biological sample -  
 XX  
 XX Example 2; Page 66-69; 93pp; Japanese.  
 XX  
 CC The present invention describes a method for evaluating renal function.  
 CC The method comprises assaying mesgin protein in biological sample. Also  
 CC described are: (1) use of an anti-mesgin protein antibody for diagnosing  
 CC renal function; and (2) a kit for detecting mesgin protein comprising:  
 CC (a) anti-mesgin protein antibody attached to solid magnetic particles;  
 CC (b) direct or indirect fixing for the antibody to the particles; and  
 CC (c) a magnet. The process is useful for evaluating renal function and  
 CC diagnosing renal disorders by assaying mesgin protein in biological  
 CC samples (preferably urine or blood). The process is reproducible and  
 CC gives accurate results. The present sequence represents the human mesgin  
 CC protein, which is given in the exemplification of the present invention.  
 XX  
 SQ Sequence 380 AA:

Query Match 100.0%; Score 1959; DB 21; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 3.6e-169;  
 Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASLAANAARECFNLFREMDNNGNGNFFSSLSLFAALALVRLGAODDSLQIDKLLHV 60  
 DB 1 MASLAANAARECFNLFREMDNNGNGNFFSSLSLFAALALVRLGAODDSLQIDKLLHV 60  
 QY 61 NTASGYGNSNSQSGLQSLKRVFSDINASHKDYDLSIVNGLFAEKYVGFHKDYIECAEK 120  
 DB 61 NTASGYGNSNSQSGLQSLKRVFSDINASHKDYDLSIVNGLFAEKYVGFHKDYIECAEK 120  
 QY 121 LYDAKVERVDFTNHLIEDTRRNINKNVENETHGKIKNVIGEGISSAVMVLVNAVYFPGK 180  
 DB 121 LYDAKVERVDFTNHLIEDTRRNINKNVENETHGKIKNVIGEGISSAVMVLVNAVYFPGK 180  
 QY 181 WQSAFTKSETINCHFKSPKSGKAVAMHOKRKFPLSYIEDPSMKILELRYNGGINMYVL 240  
 DB 181 WQSAFTKSETINCHFKSPKSGKAVAMHOKRKFPLSYIEDPSMKILELRYNGGINMYVL 240  
 QY 241 LPENDLSEIENKLFONLMEWTNPRRMTSKYVEVFFPOFKIEKNEMQYLRALGLKIDIF 300  
 DB 241 LPENDLSEIENKLFONLMEWTNPRRMTSKYVEVFFPOFKIEKNEMQYLRALGLKIDIF 300  
 QY 301 DESKADLSGIASGRLYISRMHKSYLEVTEEGTEATATGSNIYERKOLPOSTLFRADHP 360  
 DB 301 DESKADLSGIASGRLYISRMHKSYLEVTEEGTEATATGSNIYERKOLPOSTLFRADHP 360  
 QY 361 FLFVIRKDDIILFSGKVSACP 380  
 DB 361 FLFVIRKDDIILFSGKVSACP 380

RESULT 4  
 AAB83075  
 ID AAB83075 standard: Protein: 380 AA.  
 XX AAB83075;  
 AC  
 XX

DT 10-JUL-2001 (first entry)  
 XX  
 XX Human mesgin protein.  
 XX  
 KW Human; mesgin; mesangial cell proliferative nephritis; nephrotropic;  
 KW transgenic mouse; glomerular disease; animal model; drug screening.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200124628-A1.  
 XX  
 XX 12-Apr-2001.  
 PD  
 XX  
 XX 06-OCT-2000; 2000WO-JP06988.  
 PE  
 XX  
 XX 06-OCT-1999; 99JP-0285736.  
 PR  
 XX  
 PA (KURO/) KUROKAWA K.  
 PA (MIYA/) MIYATA T.  
 XX  
 XX Miyata T.  
 PI  
 XX  
 XX WPI: 2001-300136/31.  
 DR  
 XX N-PSDB: AAF82438.  
 XX  
 PT Mouse model for mesangial cell proliferative nephritis for development  
 PT and screening of new treatments -  
 XX  
 XX Example 4; Page 44-46; 62pp; Japanese.  
 PS  
 CC The present sequence is human mesgin. The human mesgin coding  
 CC sequence may be introduced into a mouse to produce an animal model of  
 CC mesangial cell proliferative nephritis. The symptoms include  
 CC enlargement of the mesangial base region, sedimentation of an immune  
 CC complex and an increase in mesangial cells. The animal model is useful  
 CC for analysing the pathology of chronic glomerular diseases and for  
 CC screening compositions for prevention and treatment of the diseases.  
 CC Highly uniform models can be made easily and in large numbers using  
 CC this method.  
 XX  
 SQ Sequence 380 AA:

Query Match 100.0%; Score 1959; DB 22; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 3.6e-169;  
 Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASLAANAARECFNLFREMDNNGNGNFFSSLSLFAALALVRLGAODDSLQIDKLLHV 60  
 DB 1 MASLAANAARECFNLFREMDNNGNGNFFSSLSLFAALALVRLGAODDSLQIDKLLHV 60  
 QY 61 NTASGYGNSNSQSGLQSLKRVFSDINASHKDYDLSIVNGLFAEKYVGFHKDYIECAEK 120  
 DB 61 NTASGYGNSNSQSGLQSLKRVFSDINASHKDYDLSIVNGLFAEKYVGFHKDYIECAEK 120  
 QY 121 LYDAKVERVDFTNHLIEDTRRNINKNVENETHGKIKNVIGEGISSAVMVLVNAVYFPGK 180  
 DB 121 LYDAKVERVDFTNHLIEDTRRNINKNVENETHGKIKNVIGEGISSAVMVLVNAVYFPGK 180  
 QY 181 WQSAFTKSETINCHFKSPKSGKAVAMHOKRKFPLSYIEDPSMKILELRYNGGINMYVL 240  
 DB 181 WQSAFTKSETINCHFKSPKSGKAVAMHOKRKFPLSYIEDPSMKILELRYNGGINMYVL 240  
 QY 241 LPENDLSEIENKLFONLMEWTNPRRMTSKYVEVFFPOFKIEKNEMQYLRALGLKIDIF 300  
 DB 241 LPENDLSEIENKLFONLMEWTNPRRMTSKYVEVFFPOFKIEKNEMQYLRALGLKIDIF 300  
 QY 301 DESKADLSGIASGRLYISRMHKSYLEVTEEGTEATATGSNIYERKOLPOSTLFRADHP 360  
 DB 301 DESKADLSGIASGRLYISRMHKSYLEVTEEGTEATATGSNIYERKOLPOSTLFRADHP 360  
 QY 361 FLFVIRKDDIILFSGKVSACP 380  
 DB 361 FLFVIRKDDIILFSGKVSACP 380

RESULT 5  
AAC64286  
ID AAG64286 standard; Protein; 380 AA.

XX AAG64286;

XX 21-SEP-2001 (first entry)

XX Rat mesgin protein.

XX Rat; mesgin; renal mesangial cell; mesangium proliferative nephritis.

XX Rattus norvegicus.

XX WO200148019-A1.

XX 05-JUL-2001.

XX 26-DEC-2000; 2000WO-JP09251.

XX 28-DEC-1999; 99JP-0373677.

XX (KURO/) KUROKAWA K.

XX (MIYA/) MIYATA T.

XX Miyata T.

XX WPI; 2001-425651/45.

XX N-PSDB; AAA48181.

XX New antibody recognizing a partial sequence of rat mesgin protein for diagnosis of mesangium proliferative nephritis

XX Disclosure; Page 54-56; 63pp; Japanese.

XX The present invention relates to a novel antibody which recognises a peptide consisting of residues 341-354 of rat mesgin protein. The present sequence is the protein sequence for rat mesgin, which was used in the present invention. Mesgin is highly expressed in renal mesangial cells and its level is elevated in mesangium proliferative nephritis. Assay of the serum or urine level using the antibody is therefore indicative of this type of disorder.

XX Sequence 380 AA;

XX Query Match 76.0%; Score 1488; DB 22; Length 380;

XX Best Local Similarity 73.7%; Pred. No. 2, 1e-126; Mismatches 43; Indels 0; Gaps 0;

XX Matches 280; Conservative 57; Mismatches 43; Indels 0; Gaps 0;

QY 1 MASLAANAEEFCNLFREMDNDNGNGNVFSSLSLFAALALVRLGAQDSSLQIDKLHY 60  
 DB 1 MASLAANAEEFGDLREMDSSOGNGNVFSSLSLFTALSLRLGARGCARQIDKALHF 60  
 QY 61 NTASGYGNSNSGSLQSLKRVFSDINASHKDYDLSYNGLFAEKVYGFHKDYIECAEK 120  
 DB 61 ISPSRQNSNSGSLQGLQYOLKRVLADINSHKDYELSIANGVFAEKVDFHRSKYMCAEN 120  
 QY 121 LDYAKERYDFTNHLDETRRNINKWVENETHGKIKRVIGEGGISSAVVAVLVNAVYFKGK 180  
 DB 121 LYNAKERYDFTNDIDETRFKINKWLENETHGKIKRVIGDSSLSASSAVVAVLVNAVYFKGK 180  
 QY 181 WSAFTKSETINCHFEKSPKSGKAVAMMHOERFNLVIEDPSMKILLELYNGGIMYVL 240  
 DB 181 WSAFTKSDTSLCHFEKSPGPKAVAMMHOERFNLSTIOEPPMQLLELYHGIGISMYYIM 240  
 QY 241 LPENDLSELENKLTFLONLMEWTNPRMTSKYVEVFPPOFKIEKNYMKOYLALGKIDF 300  
 DB 241 LPEDDLSEIESKLSFQNLMDWTNSRKMSQYVAVFLPQFKIEKDYEMRSLKSVGLEDF 300  
 QY 301 DESKADLSGIASGRLYISRMAMKSYIEVTEGSTETATAGSNIYERKQLPOSTLFRADP 360  
 DB 301 DESKADLSGIASGRLYISRMAMKSYIEVTEGSTETATAGSNIYERKQLPOSTLFRADP 360

DB 301 VESRADLSGIASGRLYISKLMHKSLEIVSEGSTETATAGSNIYERKQLPESTVFRADP 360  
 QY 361 FLFVIRKDDIILFSGKRVSCP 380  
 DB 361 FLFVIRKNGIILFTGKRVSCP 380

RESULT 6  
AAV08255  
ID AAV08255 standard; Protein; 380 AA.

XX AAV08255;

XX 14-JUL-1999 (first entry)

XX Rat mesgin protein.

XX Mesgin; mesangial cell; treatment; diagnosis; disease; Iga nephropathy;

XX human; rat; murine.

XX Rattus rattus.

XX WO9915652-A1.

XX 01-APR-1999.

XX 22-SEP-1998; 98WO-JP04269.

XX 22-SEP-1997; 97JP-0275302.

XX (KURO/) KUROKAWA K.

XX (MIYA/) MIYATA T.

XX Miyata T.

XX WPI; 1999-276983/23.

XX N-PSDB; AAX56712.

XX Mesgin protein expressed specifically in mesangial cells  
 XX Claim 1; Page 69-72; 100pp; Japanese.  
 XX This invention describes the isolation of novel mesgin nucleic acid and proteins from human, rat and mouse tissue. This protein is expressed specifically in mesangial cells. The products of the invention are useful for the treatment and diagnosis of diseases involving mesangial cells, such as Iga nephropathy.

XX Sequence 380 AA;

XX Query Match 75.4%; Score 1477; DB 20; Length 380;

XX Best Local Similarity 73.4%; Pred. No. 2e-125; Mismatches 45; Indels 0; Gaps 0;

XX Matches 279; Conservative 56; Mismatches 45; Indels 0; Gaps 0;

QY 1 MASLAANAEEFCNLFREMDNDNGNGNVFSSLSLFAALALVRLGAQDSSLQIDKLHY 60  
 DB 1 MASLAANAEEFGDLREMDSSOGNGNVFSSLSLFTALSLRLGARGCARQIDKALHF 60  
 QY 61 NTASGYGNSNSGSLQSLKRVFSDINASHKDYDLSYNGLFAEKVYGFHKDYIECAEK 120  
 DB 61 ISPSRQNSNSGSLQGLQYOLKRVLADINSHKDYELSIANGVFAEKVDFHRSKYMCAEN 120  
 QY 121 LDYAKERYDFTNHLDETRRNINKWVENETHGKIKRVIGEGGISSAVVAVLVNAVYFKGK 180  
 DB 121 LYNAKERYDFTNDIDETRFKINKWLENETHGKIKRVIGDSSLSASSAVVAVLVNAVYFKGK 180  
 QY 181 WSAFTKSETINCHFEKSPKSGKAVAMMHOERFNLVIEDPSMKILLELYNGGIMYVL 240  
 DB 181 WSAFTKSDTSLCHFEKSPGPKAVAMMHOERFNLSTIOEPPMQLLELYHGIGISMYYIM 240  
 QY 241 LPENDLSELENKLTFLONLMEWTNPRMTSKYVEVFPPOFKIEKNYMKOYLALGKIDF 300  
 DB 241 LPEDDLSEIESKLSFQNLMDWTNSRKMSQYVAVFLPQFKIEKDYEMRSLKSVGLEDF 300

```

Oy 301 DESKADISGIGSRGLYTSRMWHSXYIVTEEGTATATAGTSNIYEXOLPOSTLRADP 360
Db 301 VESADADLSGIGSRGLYVSKLMHSLSIEVSEEGTATATAGTSNIYEXOLPESTRADRP 360
Oy 361 FLFVIRKDDIILFSGKVSCP 380
Db 361 FLFVIRKNGIILFTGKVSCP 380

RESULT 7
AAB24150
ID AAB24150 standard; Protein: 380 AA.
AC AAB24150:
XX 30-JAN-2001 (first entry)
DT
DE Rat megalin protein sequence SEQ ID NO:19.
XX
XX Megalin: mesangium-predominant gene; serpin regulated; nephropathy;
KM IGA: immunoglobulin A; detection: renal function; renal disorder;
KM diagnosis: biological sample; blood; urine.
XX
OS Rattus norvegicus.
XX
FH Key Location/Qualifiers
FT Misc-difference 51
FT /note= "unspecified"
FT Misc-difference 94
FT /note= "unspecified"
XX
XX WO200057189-A1.
XX
XX 28-SEP-2000.
XX
XX 17-MAR-2000; 2000WO-JP01646.
XX
XX 19-MAR-1999; 99JP-0075305.
PR 28-OCT-1999; 99JP-0306623.
XX
XX (KUROO/) KUROKAWA K.
PA (FUSO.) FUSO PHARM IND LTD.
PA (MIYA/) MIYATA T.
XX
XX Miyata T;
PI
XX WPI: 2000-611642/58.
DR N-PSDB: AAC55238.
XX
XX Evaluating renal function comprises assaying megalin protein in
PT biological sample -
PT
XX
XX Example 2: Page 81-84; 93pp; Japanese.
XX
XX The present invention describes a method for evaluating renal function.
CC The method comprises assaying megalin protein in biological sample. Also
CC described are: (1) use of an anti-megalin protein antibody for diagnosing
CC renal function; and (2) a kit for detecting megalin protein comprising:
CC (a) anti-megalin protein antibody attached to solid magnetic particles;
CC (b) direct or indirect fixing for the antibody to the particles; and
CC (c) a magnet. The process is useful for evaluating renal function and
CC diagnosing renal disorders by assaying megalin protein in biological
CC samples (preferably urine or blood). The process is reproducible and
CC gives accurate results. The present sequence represents the rat megalin
CC protein, which is given in the exemplification of the present invention.
XX
XX Sequence 380 AA:
SQ
Query Match 75.4% Score 1477: DB 21: Length 380:
Best Local Similarity 73.4% Pred. NO. 2e-125:
Matches 279: Conservative 56: Mismatches 45: Indels 0: Gaps 0:

```

QY	1	MASIAAANAEECFNLFRMEDONONGNVEFSSLLEFALALVRLGAQDSDLSQIDKLLHV	60
Dd	1	MASIAAANAEECFDLFRMEDSSOGNGNVEFSSLSLFTLALSLIRGARDCARQIDKALHF	60
QY	61	NTASGYGNSNSOSGLOSLKRVFSDINASHKDYDLSTVNGLFAEKVYGFHKDYIECAEK	120
Dd	61	ISPSRGNSSNSQGLQYOLKRLVLADINSSHKDXLSIANGVFAEKVDFEHRKSYMECAEN	120
QY	121	LYDAKVERVPTNHLIEDTRNINKVVENETHGKIKNYGEGGISSAAVMVLVNAVYFNGK	180
Dd	121	LYNAKVRVPTNDIOETRFKINWIENETHGKIKRVGDDSSLSSAAVMVLNAVYFNGK	180
QY	181	WOSAFTRSETINCFHSPKSGCGKAVAMHOERKFNLSYIEBPSMKLLEFRNGGINVYL	240
Dd	181	WKSAPTSIDLTSCHFSPSGCGKAVAMHOERREFLSLTOEPPMOJLEQYHGGISMTIM	240
QY	241	LPENDLSIEIENKLFNFONLMEWTNPRRMYSKVEVEFPQFKIEKNEMKQYLRALGKDF	300
Dd	241	LPEDDLSEIEBKLSFONLMDWTNSRKMSQYVNVFLPQFKIEKQEMSHLSKSVLEDF	300
QY	301	DESKADLSGASGRLYISRMHKSYLEVTEEGTEATAGTSNIVEKOLPOSTLEFRADHP	360
Dd	301	VESRADLSGASGRLYYSKLMHRSLSIEVEEGTEATAGTSNIVEKLLPESTVERADRP	360
QY	361	FLFYIRKDDIILFSGKVS CP 380	
Dd	361	FLFYIRKNGLILFTGKVS CP 380	
RESULT 8			
ID	AAB83076	standard; Protein: 380 AA.	
XX	AAB83076:		
AC			
XX			
DT	10-JUL-2001	(first entry)	
XX			
DE		Rat megin protein.	
XX			
KW		Rat; mesangial cell proliferative nephritis; nephrotropic;	
KW		transgenic mouse; glomerular disease; animal model; drug screening.	
XX			
OS		Rattus norvegicus.	
XX			
PN	WO200124628-A1.		
PD	12-APR-2001.		
XX			
PF	06-OCT-2000; 2000WO-JP06988.		
XX			
PR	06-OCT-1999; 99JP-0285736.		
PA	(KUROO/) KUROKAWA K.		
PA	(MIYATA/) MIYATA T.		
PI	Miyata T;		
XX			
DR	WPI: 2001-300136/31.		
DR	N-PSDB; AAF82439.		
XX			
PT	Mouse model for mesangial cell proliferative nephritis for development		
PT	and screening of new treatments		
XX			
PS	Disclosure; Page 48-50; 62pp; Japanese.		
XX			
CC	The present sequence is rat megin. The human megin coding		
CC	sequence may be introduced into a mouse to produce an animal model of		
CC	mesangial cell proliferative nephritis. The symptoms include		
CC	enlargement of the mesangial base region, sedimentation of an immune		
CC	complex and an increase in mesangial cells. The animal model is useful		
CC	for analysing the pathology of chronic glomerular diseases and for		
CC	screening compositions for prevention and treatment of the diseases.		
CC	Highly uniform models can be made easily and in large numbers using		

CC this method.  
 XX Sequence 380 AA;  
 SQ  
 Query Match 75.4%; Score 1477; DB 22; Length 380;  
 Best Local Similarity 73.4%; Pred. No. 2e-125; Mismatches 45; Indels 0; Gaps 0;  
 Matches 279; Conservative 56; Mismatches 45; Indels 0; Gaps 0;  
 QY 1 MASLAANAEEFCNLFREKMDNONGNNGVNFSSLSLFAALALVRLGADDSLSQIDKLHV 60  
 1 MASLAANAEEFGDLFREMDSSQNGNVFSSLSLFTALIRLGAGDCXQRODKALHF 60  
 DB 61 NTASGVYSSNSQSGLOSQIKRVPFSDINASHKDYDLSIVGLFAEKVYGFHKDYIECAEK 120  
 61 ISSRQGNSSNSQIGLOYLKRVLADINSSHKDKLSIANGVFAEKVDFHKSYMECAEN 120  
 QY 121 LYPAKVERVDFTNLEDTFRNRINKVVENETHGKIKNYIGGJSSSAVWLVNAVYFKGK 180  
 121 LYNAKVERVDFTNDIOETREKINKWVENETHGKIKKAVLGSSSSSAVWLVNAVYFKGK 180  
 QY 181 WQSAFTKSETINCHFKSPKSGRAVAMHOREKFNLSVIEDPSMKILELRYNGSINMYVL 240  
 181 WSAFTKSDTLSCHFHRSQPGKAVNMHOREFNLSIOEPWQILELQYHGISMVIM 240  
 DB 241 LPEVDLSEIKNKLTFONLMEWTNPRRTSKYVEYFPOFKIEKNYEMKQYLRLALGKDIF 300  
 241 LPDDLSLEISKLSLFONLMDWTNSRKMSQYVNVFLPQFKIEKDYEMRSHLSGVLEDF 300  
 QY 301 DESKADLSIASGGRLEYISHMHKSYIEVTEEGEATATGNSNIVEKQLPOSTLERADHP 360  
 301 VESRADLSIASGGRLEYISHMHKSYIEVSEEGEATATGNSNIVEKQLPESTVFRADRP 360  
 DB 361 FLEVIKRDIIILFSGKVS 380  
 361 FLEVIKRGIIILFTGKVS 380  
 QY  
 DB  
 RESULT 9  
 AA08256  
 ID AAY08256 standard; Protein: 368 AA.  
 XX  
 AC AAY08256;  
 XX  
 DT 14-JUL-1999 (first entry)  
 XX  
 DE Mouse megin protein.  
 XX  
 KW Megin; mesangial cell; treatment; diagnosis; disease; Iga nephropathy;  
 KM human; rat; murine.  
 XX  
 OS Mus musculus.  
 XX  
 FH Key Location/Qualifiers  
 FT Protein 1..368  
 XX /note="partial sequence"  
 FT  
 XX W09915652-A1.  
 XX  
 PD 01-APR-1999.  
 XX  
 PF 22-SEP-1998; 98WO-JP04269.  
 XX  
 PR 22-SEP-1997; 97JP-0275302.  
 XX  
 PA (KURO/) KUROKAWA K.  
 PA (MIYA/) MIYATA T.  
 XX  
 PI Miyata T;  
 XX  
 DR WPI: 1999-276983/23.  
 DR N-PSDB: AAX56714.  
 XX  
 PT Megin protein expressed specifically in mesangial cells

XX  
 PS Claim 1; Page 76-79; 100pp; Japanese.  
 XX  
 CC This invention describes the isolation of novel megin nucleic acid and  
 CC proteins from human, rat and mouse tissue. This protein is expressed  
 CC specifically in mesangial cells. The products of the invention are  
 CC useful for the treatment and diagnosis of diseases involving mesangial  
 CC cells, such as Iga nephropathy.  
 XX  
 SQ Sequence 368 AA;  
 Query Match 74.4%; Score 1458; DB 20; Length 368;  
 Best Local Similarity 73.4%; Pred. No. 1e-123;  
 Matches 270; Conservative 53; Mismatches 45; Indels 0; Gaps 0;  
 QY 13 FNLFREKMDNONGNNGVNFSSLSLFAALALVRLGADDSLSQIDKLHVNTASGVYSSNS 72  
 1 FDLFREMDSSQNGNVFSSLSLFTALTIRLGAGDCARQIDKALHFNIPSRQGNSSNN 60  
 DB 73 QSGLOQLKRVFSDINASHKDYDLSIVGLFAEKVYGFHKDYIECAEKLYDAKVERVDFT 132  
 61 QPGLQYQKRVLADINSSHKDYELSTATGYFAEKVYDFHKNYIECAENLYNAKVERVDFT 120  
 QY 133 NHEEDTRNRINKVVENETHGKIKNYIGEGCISSAVWLVNAVYFKGKQSAFTKSETIN 192  
 133 NHEEDTRNRINKVVENETHGKIKNYIGEGCISSAVWLVNAVYFKGKQSAFTKSETIN 192  
 DB 121 NDVQDTRFKINKWVENETHGKIKRVLDSSLSA VNAVYFKGKMSAFTKSDTL 180  
 QY 193 CHFSPKSGRAVAMHOREKFNLSVIEDPSMKILELRYNGSINMYVLPEVDLSEIKN 252  
 181 CRFRSPCPGKAVNMHOREFNLSIQPWOYLELQYHGISMVIMPEVDLCEIESK 240  
 DB 241 LSFQNLMDWTNRRKMSQYVNVFLPQFKIEKNYEMTHHLSGLKIDFDESSADLSIAS 300  
 QY 253 LTFONLMEWTNPRRTSKYVEYFPOFKIEKNYEMKQYLRLALGKIDFDESKADLSIAS 312  
 241 LSFQNLMDWTNRRKMSQYVNVFLPQFKIEKNYEMTHHLSGLKIDFDESSADLSIAS 300  
 DB 313 GGRLYISRMHKSIEVTEEGEATATGNSNIVEKQLPOSTLERADHPFLVIRKDDIIL 372  
 301 GGRLYYSKLMHKSIEVSEEGEATATGNSNIVEKQLPESTVFRADRPFLVIRKDDIIL 360  
 QY 373 FSGKVS 380  
 361 FTGKVS 368  
 DB  
 RESULT 10  
 AAB24151  
 ID AAB24151 standard; Protein: 368 AA.  
 XX  
 AC AAB24151;  
 XX  
 DT 30-JAN-2001 (first entry)  
 XX  
 DE Mouse megin protein sequence SEQ ID NO:21.  
 XX  
 KW Megin; mesangium-predominant gene; serpin regulated; nephropathy;  
 KM Iga; immunoglobulin A; detection; renal function; renal disorder;  
 XX diagnosis; biological sample; blood; urine.  
 XX  
 OS Mus musculus.  
 XX  
 PN W0200057189-A1.  
 XX  
 PD 28-SEP-2000.  
 XX  
 PF 17-MAR-2000; 2000WO-JP01646.  
 XX  
 PR 19-MAR-1999; 99JP-0075305.  
 PR 28-OCT-1999; 99JP-0306623.  
 XX  
 PA (KURO/) KUROKAWA K.  
 PA (FUSO/) FUSO PHARM IND LTD.  
 PA (MIYA/) MIYATA T.  
 XX



PI Miyata T;  
 XX WPI: 2000-611642/58.  
 DR N-PSDB: AAC55239.  
 XX  
 PT Evaluating renal function comprises assaying mesgin protein in  
 PT biological sample  
 XX  
 PS Disclosure: Page 89-91; 93pp; Japanese.  
 XX  
 CC The present invention describes a method for evaluating renal function.  
 CC The method comprises assaying mesgin protein in biological sample. Also  
 CC described are: (1) use of an anti-mesgin protein antibody for diagnosing  
 CC renal function; and (2) a kit for detecting mesgin protein comprising:  
 CC (a) anti-mesgin protein antibody attached to solid magnetic particles;  
 CC (b) direct or indirect fixing for the antibody to the particles; and  
 CC (c) a magnet. The process is useful for evaluating renal function and  
 CC diagnosing renal disorders by assaying mesgin protein in biological  
 CC samples (preferably urine or blood). The process is reproducible and  
 CC gives accurate results. The present sequence represents the mouse mesgin  
 CC protein, which is given in the exemplification of the present invention.  
 CC  
 SQ Sequence 368 AA:

Query Match 74.4%; Score 1458; DB 21; Length 368;  
 Best Local Similarity 73.4%; Pred. No. 1e-123;  
 Matches 270; Conservative 53; Mismatches 45; Indels 0; Gaps 0;

OY 13 FNLFRMDNDGNGVNFSSLSLFAALALVRLGADDSLSQIDKLHVNATAGYGNSSNS 72  
 DB 1 FDLFRMDNSGNGVNFSSLSIFALTLRLRGAGDCARQIDKALHFNIPSRGSSNN 60

OY 73 QSGLOSQLRVFSIDINASHKDYLSIVNGLFAEKYVGFHKDYIECAEKLVDKVERVDFT 132  
 DB 61 QPGLOQLKRVLADINSSHKDYELISATGVFAEKYVDFHKNYIECAENLYNAKVERVDFT 120

OY 133 NHELETRRNINKWVENETHGKIKNYIGEGISSAVMVLVNAVYFKGKQSAFTKSETIN 192  
 DB 121 NDVQDTRRKINKWENETHGKIKNYIGEGISSAVMVLVNAVYFKGKQSAFTKTDLS 180

OY 193 CHEFSPKSGKAVAMHOKRFNLSVIDPSSKILIELRYNGGINNYVLLPENDLSEIENK 252  
 DB 181 CRFSPKCPGKAVNMHOKRFNLSVIDPSSKILIELRYNGGINNYVLLPENDLSEIESK 240

OY 253 LTFQNLMEWTNPRMTSKYVEVEFPQFIENKNEYKQYLRLALGLDIDESSADLSGIAS 312  
 DB 241 LSFQNLMDWTNRKMKSQYVAVFLPQFIENKNEYTHLKSGLDIDESSADLSGIAS 300

OY 313 GGRLYISRMHMKSYLEVTEEGTEATAGSNIVKQLPOSTLFRADHPFLFYIRKDDITL 372  
 DB 301 GGRLYVSKLMHKSFLVSEEGTEATAENNIVKQLPESTVFRADRPFLFYIKKNDITL 360

OY 373 FSGKVSCLP 380  
 DB 361 FTGKVSCLP 368

RESULT 11  
 AAB83077  
 ID AAB83077 standard; Protein: 368 AA.

XX AAB83077;  
 XX  
 DT 10-JUL-2001 (first entry)

XX Murine mesgin protein.  
 XX  
 KW Mouse; mesangial cell proliferative nephritis; nephrotropic;  
 XX transgenic mouse; glomerular disease; animal model; drug screening.

OS Mus musculus.  
 XX  
 PN WO200124628-A1.

XX 12-APR-2001.  
 PD  
 XX 06-OCT-2000; 2000WO-JP06988.  
 PF  
 XX 06-OCT-1999; 99JP-0285736.  
 PR  
 XX (KUROO) KUROKAWA K.  
 PA (MIYA/) MIYATA T.  
 PA

PI Miyata T;  
 XX WPI: 2001-300136/31.  
 DR N-PSDB: AAF82440.  
 XX

PT Mouse model for mesangial cell proliferative nephritis for development  
 PT and screening of new treatments  
 XX  
 PS Disclosure: Page 52-53; 62pp; Japanese.

CC The present sequence is murine mesgin protein. The human mesgin coding  
 CC sequence may be introduced into a mouse to produce an animal model of  
 CC mesangial cell proliferative nephritis. The symptoms include  
 CC enlargement of the mesangial base region, sedimentation of an immune  
 CC complex and an increase in mesangial cells. The animal model is useful  
 CC for analysing the pathology of chronic glomerular diseases and for  
 CC screening compositions for prevention and treatment of the diseases.  
 CC Highly uniform models can be made easily and in large numbers using  
 CC this method.

SQ Sequence 368 AA:

Query Match 74.4%; Score 1458; DB 22; Length 368;  
 Best Local Similarity 73.4%; Pred. No. 1e-123;  
 Matches 270; Conservative 53; Mismatches 45; Indels 0; Gaps 0;

OY 13 FNLFRMDNDGNGVNFSSLSLFAALALVRLGADDSLSQIDKLHVNATAGYGNSSNS 72  
 DB 1 FDLFRMDNSGNGVNFSSLSIFALTLRLRGAGDCARQIDKALHFNIPSRGSSNN 60

OY 73 QSGLOSQLRVFSIDINASHKDYLSIVNGLFAEKYVGFHKDYIECAEKLVDKVERVDFT 132  
 DB 61 QPGLOQLKRVLADINSSHKDYELISATGVFAEKYVDFHKNYIECAENLYNAKVERVDFT 120

OY 133 NHELETRRNINKWVENETHGKIKNYIGEGISSAVMVLVNAVYFKGKQSAFTKSETIN 192  
 DB 121 NDVQDTRRKINKWENETHGKIKNYIGEGISSAVMVLVNAVYFKGKQSAFTKTDLS 180

OY 193 CHEFSPKSGKAVAMHOKRFNLSVIDPSSKILIELRYNGGINNYVLLPENDLSEIENK 252  
 DB 181 CRFSPKCPGKAVNMHOKRFNLSVIDPSSKILIELRYNGGINNYVLLPENDLSEIESK 240

OY 253 LTFQNLMEWTNPRMTSKYVEVEFPQFIENKNEYKQYLRLALGLDIDESSADLSGIAS 312  
 DB 241 LSFQNLMDWTNRKMKSQYVAVFLPQFIENKNEYTHLKSGLDIDESSADLSGIAS 300

OY 313 GGRLYISRMHMKSYLEVTEEGTEATAAGSNIVKQLPOSTLFRADHPFLFYIRKDDITL 372  
 DB 301 GGRLYVSKLMHKSFLVSEEGTEATAENNIVKQLPESTVFRADRPFLFYIKKNDITL 360

OY 373 FSGKVSCLP 380  
 DB 361 FTGKVSCLP 368

RESULT 12  
 AAY15155  
 ID AAY15155 standard; Protein: 379 AA.

XX AAY15155;  
 XX  
 DT 07-FEB-2000 (first entry)

DE Human prostate serine protease inhibitor (PROSTAPIN).  
 XX  
 KW PROSTAPIN; serine protease inhibitor; serpin family; prostate; human;  
 KM prostate-specific tumor suppressor; apoptosis-inducer; diagnosis;  
 KM apoptosis-modulator; prognosis; treatment; prostate cancer;  
 KM metastatic prostate cancer.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Inhibitory-site 330..347  
 FT /label= "Serine-protease-inhibitory-domain  
 FT /note= "Also known as reactive-site loop (RSL) and  
 FT contains a hinge region and a variable region. RSL  
 FT binds to protease active site during inhibition"  
 FT 330..338  
 FT Region /label= "RSL-hinge region  
 FT /note= "confers stability to the serpin-protease  
 FT complex and is highly conserved among the serpins"  
 FT 344..345  
 FT Cleavage-site /note= "included in the RSL variable region. PROSTAPIN  
 FT undergoes cleavage at this site during inhibition"  
 XX  
 PN W0958560-A2.  
 XX  
 PD 18-NOV-1999.  
 XX  
 PF 31-MAR-1999; 99WO-US07123.  
 XX  
 PR 31-MAR-1998; 98US-0080167.  
 PR 15-MAY-1998; 98US-0085720.  
 XX  
 PA (UROG-) UROGENESYS INC.  
 PA (AFAR/) AFAR D E.  
 PA (HUBE/) HUBERT R S.  
 PA (LEON/) LEONG K.  
 PA (RAIT/) RAITANO A B.  
 PA (SAFF/) SAFFRAN D C.  
 XX  
 PI Afar DE, Hubert RS, Leong K, Raitano AB, Saffran DC;  
 DR WPI: 2000-039068/03.  
 DR N-PSDB; AA229049.  
 XX  
 PT New isolated prostapin gene, used to develop products for, e.g.  
 PT treating and determining susceptibility to prostate cancer  
 XX  
 PS Claim 1; Fig 1; 63pp; English.  
 XX  
 CC The present sequence is that of a PROSTAPIN protein which is a member of  
 CC serine protease inhibitor (serpin) family and expressed in normal  
 CC prostate and locally confined prostate cancer. The sequence is derived  
 CC from overlapping sequences of human cDNA clones 11P26 and 103. PROSTAPIN  
 CC contains a reactive-site loop characteristic of serpin family in  
 CC the carboxy-terminal region. The protein may function as a prostate-  
 CC specific tumour suppressor, apoptosis-inducer or apoptosis-modulator.  
 CC The sequence is useful in diagnosing, prognosing, monitoring and  
 CC treating prostate cancer, particularly advanced stage and metastatic  
 CC prostate cancer.  
 CC  
 SQ Sequence 379 AA;  
 Query Match 40.7%; Score 796.5; DB 21; Length 379;  
 Best Local Similarity 42.6%; Pred. No. 1.2e-63;  
 Matches 168; Conservative 77; Mismatches 120; Indels 29; Gaps 8;

QY 114 YIECAEKLVDKAKYERVDFTNHLIEDTRRNINKWVENETHGKIKNVIGEGGISSAVMYLVN 173  
 Db 121 YLSCSEKWKYARLQTVDFEOSTETREKMINAVENKNGKAVANLFGKSTIDPSSVMVLVN 180  
 QY 174 AVYFKGKWSAFKSEITNHFSPKSGR--AVAMMHQRKNLSVIEDEPSKILELRY 231  
 Db 181 TIFYFGGRONKF-----QGNVTVEMVYQIGTFKLAFFKEPQWVLELPY 225  
 QY 232 -NGINMYVLLPEN--DLSEIENKLPFQNMEMTNPBRMYSKYVEFFPOFKIEKNEMK 288  
 Db 226 VNNKLSMILLPFGIANLKEQKNGTTHEMTSSNMHREVEVHLLPFKLEIKLEYLN 285  
 QY 289 QYLRALGLKIDFESKADLSGIASGRLYISRMHMSYIEVEEGTEATPAATGSNIVEQ 348  
 Db 286 SLKPLGVTDLFNQVKADLSGMSPTKGLYSKAIHKSXYLDVSEGEFAAATGDSIAVKS 345  
 QY 349 LPOSTLFRADHPFLFYVR--KDIILFSGVSCP 380  
 Db 346 LPMRAQFKANHPFLFYRHTHTNTVILFCGLASP 379  
 RESULT 13  
 AAM47207 standard; Protein; 395 AA.  
 ID AAM47207;  
 AC AAM47207;  
 XX  
 DT 12-FEB-2002 (first entry)  
 XX  
 DE Human NOVI protein.  
 XX  
 KW Human: NOVI; metabolic disorder; neurodegenerative disorder;  
 KW immune disorder; haematopoietic disorder; developmental disease; cancer;  
 KW retinal disease; feeding disorder; vaccine; infection; gene therapy;  
 KW neurological disorder; psychotic disorder; G-protein coupled receptor;  
 KW cytosolic; antidiabetic; virucide; neuroprotective; noctropic;  
 KW analgesic; antidepressant; antimalarial; anticonvulsant; neuroleptic;  
 KW antiallergic; antinflammatory; anorectic; antiarthritic;  
 KW antiparasitic; antithrombotic; antibacterial; fungicide;  
 KW osteopathic; protozoacide; antilucer; hypertensive; hypotensive;  
 KW antifertility; vulnery; nephrotropic; antilipemic; leupin;  
 KW chromosome 18.  
 XX  
 OS Homo sapiens.  
 OS  
 XX  
 PN W0200174851-A2.  
 PD 11-OCT-2001.  
 PD  
 PF 30-MAR-2001; 2001WO-US10039.  
 XX  
 PR 30-MAR-2000; 2000US-193205P.  
 PR 30-MAR-2000; 2000US-193339P.  
 PR 05-APR-2000; 2000US-195343P.  
 PR 06-APR-2000; 2000US-195005P.  
 PR 06-APR-2000; 2000US-195088P.  
 PR 10-APR-2000; 2000US-195792P.  
 PR 11-APR-2000; 2000US-196556P.  
 PR 13-APR-2000; 2000US-197081P.  
 PR 14-APR-2000; 2000US-197087P.  
 PR 14-APR-2000; 2000US-197525P.  
 PR 29-MAR-2001; 2001US-0823187.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Meijnder K, Spaderna SK, Taupier RJ, Padigaru M, Burgess CE;  
 PI Shimkets RA, Spytek KA, Liu X, Paturajan M, Gusev VY;  
 XX  
 DR WPI: 2001-626379/72.  
 DR N-PSDB; ABA01980.  
 XX  
 PT New G protein-coupled receptor related polypeptides and polynucleotides

PT for diagnosis, prevention and treatment of metabolic,  
PT neurodegenerative, retinal, immune, hematopoietic disorders, diabetes,  
PT obesity and infections -  
XX  
Claim 1; Page 9; 194pp; English.  
PS  
XX

The present invention provides the protein and coding sequences of novel human G-protein coupled receptors, designated NOV1, NOV2, NOV3, NOV4, NOV5a, NOV5b, NOV6b, NOV7, NOV8a, NOV8b, NOV9 and NOV10. These can be used in the treatment of NOVX-related diseases, including cancer, metabolic, neurodegenerative, immune, hematopoietic, developmental, retinal, feeding, neurological and psychiatric diseases and disorders and infections. The present sequence is the NOV1 protein, the gene for which is found on chromosome 18. The NOV1 protein shares homology with the leupin protein.

**SQ Sequence 395 AA;**

Query Match	39.48;	Score 771.5;	DB 22;	Length 395;
-------------	--------	--------------	--------	-------------

Matches 159; Conservative 89; Mismatches 132

Matches 159; Conservative 89; Mismatches 132; Indels 15; Gaps 5;

[illegible]

RESULT 14  
ABB53280  
ID ABB53280 standard; Protein; 394 AA.

AC ABB53280;

DT 12-FEB-2002 (first entry)

Human polypeptide #20.

KW neurotropic; neuroprotective; anticonvulsant; antidepressant;  
 KW neuroleptic; tranquiliser; antihypertensive; cardiant; antidiabetic;  
 KW antiinflammatory; antihypaemic; hepatotropic; virucide; antidiabetic;  
 KW nephrotropic; anorectic; cytosstatic; vaccine; neurological disease;  
 KW cardiovascular disease; respiratory disease; liver disease;  
 KW renal disease; skeletal muscle disease; gastrointestinal disease;  
 KW placental disease; testicular cancer; male fertility; pancreatic disease

OS Homo sapiens.  
XX  
PN WO200181363-A1.  
XX

PD	01-NOV-2001.
XX	
PF	26-APR-2001; 2001WO-US13360.

PR 27-APR-2000; 2000US-199963P  
PR 11-MAY-2000; 2000US-203336P  
PR 25-MAY-2000; 2000US-207087P  
PR 26-MAY-2000; 2000US-207546P

PA (SMIK ) SMITHKLINE BEECHAM CORP  
PA (SMIK ) SMITHKLINE BEECHAM PLC.

PI Agarwal P, Murdock PR, Rizvi SK, Smith RF, Xiang Z, Kabnick KS,  
PI Lal Y, Xie Q;

DR WPI; 2002-041392/05.  
DR N-PSDB; ABA90345.

pt Novel polypeptides and polynucleotides useful as a vaccine for  
pt preventing and treating diseases associated the polypeptide, e.g.  
pt Alzheimer's disease, dyslipidemia, obesity, diabetes, infertility  
pt asthma, amnesias -

PS Claim 1; page 90-91; 116pp; English

The invention relates to an isolated polypeptide comprising a 277, 480, 583, 561, 628, 424, 638, 229, 310, 841, 241, 369, 382, 185, 586, 1026, 844, 782, 262, 394, 471, 485, 286, 533, 459, 350, 619, 490, 462, 253, 784, 252, 593, 472, 607, 781, 640, 686 or 154 amino acid sequence as given in the specification. The polypeptides, modulators of the polypeptides and antibodies against the polypeptides are useful for treating diseases such as neurological and psychiatric diseases including Alzheimer's, paraspinalcular palsy, Huntington's disease, myotonic dystrophy, anorexia and depression; cardiovascular diseases including congestive heart failure, Hodgkin's disease and myocardial infarction; respiratory diseases including asthma, chronic obstructive pulmonary disease, cystic fibrosis and adult respiratory distress syndrome; liver diseases including hypercholesterolaemia, cirrhosis, viral and nonviral hepatitis, Type II diabetes mellitus, and impaired glucose tolerance, renal disease including renal failure, acute tubular necrosis and glomerulonephritis; skeletal muscle diseases including Eulenberg's disease, hypoglycaemia and obesity; gastrointestinal diseases including myotonia congenita and intestinal obstruction; lymph diseases including lymphogranulocystia; diseases of placenta including chorioarctoma; diseases of testes including testicular cancer, male reproductive diseases including low testosterone and male infertility; and disease of pancreas including diabetic ketoacidosis, Type 1 and 2 diabetes and obesity. The present sequence is a polypeptide of the invention.

SQ	Sequence	394	AA
----	----------	-----	----

Query Match	39.38;	Score 769;	DB 23;	Length 394;
-------------	--------	------------	--------	-------------

Matches 158; Conservative 88; Mismatches 134; Indels 14; Gaps 5;

Matches 158; Conservative 88; Mismatches 134; Indels 14; Gaps 5;

```

QY 1 MASLAAANAEEFCNLFREMDNDQNGNVPFSSLSLEPAALALVRLGQODDSLSIDKLLHY 60
Db 1 MDLSLVANKKFFEDLFQELGKDDRRHNIFSPISLISLAAGMVLRLGARSQSAHODEVLHP 60
QY 61 NTAS-----GYGSSNSQSGLOS-QLKRVEFSDINASHKRDLSIVNGLFEAKVYGFHKDYI 115
Db 61 NKTTEPRLDQOASLNNESSLVSCYFCQGLLSKDRITDTDTLTSLANLVLGGEPPIQCEYL 120
QY 116 ECAEKLIDAKVERVDFTNHL EDTRRNINKWVENETHGKTKNYIGEGSISSAAMVLVNAV 175
Db 121 DGVIEFYHTTIESVDFOQKNPEKSRQEIINWVEQSGQGIKELFSKAIINAEVLTVLVNAV 180
QY 176 YKRGKMSAFKTSERTINCHFSRPGKCGRAVAMMHQERKFNLSIYEDPSKMTIELRY-NGS 234
Db 181 YKRAHNEYFDHENTYDAPFLCLNANENKSVKMMTQGLGRIGIEEVKAOIEMRTTRCK 240
QY 235 INMYVLLPND-----LSEIENKLTQNLMEWTNFRMTSKYVEVEFPQFIKENYEMK 288

```

Db	241	ISMFLVLPSSKDNLKGLELERTKITYEKAMVAMSSSENNSSESVLSLSPFTLEDSYDLN	300
Qy	289	QYLRALGLKDIDESKADLSGTAAGRLYISRMHKSYLEVTEEGTEATPAATGSNIYEQ	348
Db	301	SLIDMGITDIDEDERADLTGISPSFNLYLSIKHTFVEVDENGTOAAATGAIVSERS	360
Qy	349	LPQSTLFRADHPELFYIR--KDDILIFSGKVSQCP	380
Db	361	LRSWYEFNANHPELFFIRINKTQITLFTGRVCSF	394
RESULT	15		
ID	AAR25276	standard; Protein; 390 AA.	
AC	AAR25276;		
DT	08-JAN-1993	(first entry)	
XX	SCC antigen.		
XX	Squamous cell carcinoma; cancer; immunogen; monoclonal antibody;		
KW	diagnosis; serine protease inhibitor.		
OS	Synthetic.		
FH	Key	Location/Qualifiers	
FT	Region	11..21	
FT		/label= 21-2=26	
FT		/note= "region corresp. to partially determined antigen sequences"	
FT	Region	231..237	
FT		/label= 10	
FT		/note= "region corresp. to partially determined antigen sequences"	
FT	Region	241..256	
FT		/label= 21-1	
FT		/note= "region corresp. to partially determined antigen sequences"	
FT	Region	303..325	
FT		/label= 19	
FT		/note= "region corresp. to partially determined antigen sequences"	
FT	Modified-site	65..67	
FT		/note= "glycosylation site"	
FT	Modified-site	93..95	
FT		/note= "glycosylation site"	
FT	Modified-site	171..173	
FT		/note= "glycosylation site"	
FT	Modified-site	376..378	
FT		/note= "glycosylation site"	
XX	DE4139418-A.		
XX	09-JUL-1992.		
XX	29-NOV-1991;	91DE-4139418.	
XX	30-NOV-1990;	90JP-0330155..	
XX	(DAIN-) DAINABOT CO LTD.		
PA	(DAIN-) DAINABOT KK.		
XX	Kato H, Sekiguchi K, Suminami Y, Takeda K;		
XX	WPI; 1992-235415/29.		
DR	N-PSDB; AAQ25817.		
PT	New DNA encoding squamous cell carcinoma antigen - allowing large		
PT	scale antigen prodn. for use as diagnostic reagent or immunogen,		
PT	also useful as hybridisation probe		
XX			

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:49:36 ; Search time 67.9418 Seconds  
(without alignments)  
164.563 Million cell updates/sec

Title: US-10-091-442-34

Perfect score: 1959  
Sequence: 1 MASLAANAEECFNLFREMD.....FLFVRKDIILFGKVSQP 380

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Listing first 45 summaries

Issued\_Patents\_AA: \*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/1/1aa/PCTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/1/1aa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1959	100.0	380	2	US-08-472-659-34
2	1959	100.0	380	2	US-08-474-661-34
3	1959	100.0	380	2	US-08-611-977-34
4	752	38.4	390	4	US-08-568-1478-2
5	747	38.1	390	4	US-09-266-910-3
6	745	38.0	390	4	US-09-266-910-4
7	730.5	37.3	375	1	US-08-121-714-8
8	730.5	37.3	375	1	US-08-477-108A-8
9	730.5	37.3	375	5	PCT-US93-08322-8
10	730.5	37.3	375	5	PCT-US93-08322-8
11	722.5	36.9	379	1	US-08-121-714-4
12	722.5	36.9	379	1	US-08-477-108A-4
13	722.5	36.9	379	5	PCT-US93-08322-4
14	722.5	36.9	379	5	PCT-US93-08322-4
15	715.5	36.5	391	4	US-09-123-912-110
16	715.5	36.5	391	4	US-09-123-912-110
17	715	36.5	400	4	US-09-123-912-112
18	715	36.5	400	4	US-09-123-912-112
19	715	36.5	400	4	US-09-123-912-112
20	673	34.4	382	1	US-07-768-286B-6
21	673	34.4	382	2	US-08-487-823B-3
22	673	34.4	382	2	US-08-997-040-3
23	671	34.3	392	1	US-09-203-237-3
24	664.5	33.9	415	1	US-07-768-286B-4
25	664.5	33.9	415	1	US-07-911-531-19
26	664.5	33.9	415	1	US-07-693-636A-19
27	664.5	33.9	434	1	US-07-768-286B-2
27	664.5	33.9	434	1	US-07-679-052A-15

28	662.5	33.8	415	4	US-09-026-408-4	Sequence 4, Appl1
29	662	33.8	374	1	US-08-464-148-2	Sequence 2, Appl1
30	662	33.8	374	1	US-08-385-500-2	Sequence 2, Appl1
31	662	33.8	374	1	US-08-846-784-2	Sequence 2, Appl1
32	660.5	33.7	438	1	US-07-679-052A-17	Sequence 17, Appl1
33	653.5	33.4	405	1	US-08-121-714-7	Sequence 7, Appl1
34	653.5	33.4	405	1	US-08-477-108A-7	Sequence 7, Appl1
35	653.5	33.4	405	2	US-08-477-112-7	Sequence 7, Appl1
36	653.5	33.4	405	2	PCT-US93-08322-7	Sequence 7, Appl1
37	645	32.9	375	1	US-08-121-714-5	Sequence 5, Appl1
38	645	32.9	375	1	US-08-477-108A-5	Sequence 5, Appl1
39	645	32.9	375	2	US-08-477-112-5	Sequence 5, Appl1
40	645	32.9	375	5	PCT-US93-08322-5	Sequence 5, Appl1
41	642	32.8	376	4	US-09-200-965-2	Sequence 2, Appl1
42	642	31.8	376	1	US-08-464-148-4	Sequence 4, Appl1
43	623	31.8	376	1	US-08-385-500-4	Sequence 4, Appl1
44	623	31.8	376	1	US-08-846-784-4	Sequence 4, Appl1
45	592	30.2	465	3	US-08-948-997-6	Sequence 6, Appl1

#### ALIGNMENTS

RESULT 1  
US-08-472-659-34  
; Sequence 34, Application US/08472659  
; Patent No. 5831030  
; GENERAL INFORMATION:  
; APPLICANT: TSUJIMOTO, Masafumi  
; APPLICANT: IWASA, Fuyuki  
; APPLICANT: TSURODOKA, No. 5831030uo  
; APPLICANT: NAKAZATO, Hiroshi  
; APPLICANT: MIURA, Kenju  
; APPLICANT: ISHIDA, No. 5831030uh1ro  
; APPLICANT: KURIHARA, Tatsuya  
; APPLICANT: YAMAICHI, Kozo  
; APPLICANT: YAMAGUCHI, No. 5831030om1  
; TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
; NUMBER OF SEQUENCES: 34  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: P.O. Box 1404  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/472,659  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 4-212305  
; FILING DATE: 17-JUL-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 5-067339  
; FILING DATE: 04-MAR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/091,028  
; FILING DATE: 14-JUL-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McGowan, Malcolm K.  
; REGISTRATION NUMBER: 39,300  
; REFERENCE/DOCKET NUMBER: 001560-248  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 34:  
; SEQUENCE CHARACTERISTICS:

LENGTH: 380 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-472-659-34

Query Match 100.0%; Score 1959; DB 2; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 1e-178;  
 Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASLAANAEFCNLFREMDNNGNNGVFFSSLSLPAALALVRLGAQDDSLQIDKLHY 60  
 DB 1 MASLAANAEFCNLFREMDNNGNNGVFFSSLSLPAALALVRLGAQDDSLQIDKLHY 60  
 QY 61 NTASGYGNSNSGSGLOSOLKRVFSDINASHKDYDLSYNGLEPAEKYGFHKYIECAEK 120  
 DB 61 NTASGYGNSNSGSGLOSOLKRVFSDINASHKDYDLSYNGLEPAEKYGFHKYIECAEK 120  
 QY 121 LYDAKVERVDFTNHLIEDTRRNINKWVENETHGKIKNVIGSGISSAVMLVNAVYFEGK 180  
 DB 121 LYDAKVERVDFTNHLIEDTRRNINKWVENETHGKIKNVIGSGISSAVMLVNAVYFEGK 180  
 QY 181 WQSAFTKSETINCHFKSPKCSGKAVAMHQBKFNLSYIEDPSMKILIELRYNGINMYVL 240  
 DB 181 WQSAFTKSETINCHFKSPKCSGKAVAMHQBKFNLSYIEDPSMKILIELRYNGINMYVL 240  
 QY 241 LPENDLSEIENKLTFOIEMETNPRRMTSKYVEVFFPOFKIEKNYEMKOYLRLGLKIDIF 300  
 DB 241 LPENDLSEIENKLTFOIEMETNPRRMTSKYVEVFFPOFKIEKNYEMKOYLRLGLKIDIF 300  
 QY 301 DESKADLSGIASGRLYISMMHKSYTEVEEGTEATAATGSNIVEKOLPOSTLFRADHP 360  
 DB 301 DESKADLSGIASGRLYISMMHKSYTEVEEGTEATAATGSNIVEKOLPOSTLFRADHP 360  
 QY 361 FLFVIRKDDIILFSGKVS 380  
 DB 361 FLFVIRKDDIILFSGKVS 380

## RESULT 2

US-08-474-661-34  
 Sequence 34, Application US/08474661  
 Patent No. 5874253  
 GENERAL INFORMATION:  
 APPLICANT: TSUJIMOTO, Masafumi  
 APPLICANT: IWASA, Fuyuki  
 APPLICANT: TSUBUOKA, No. 5874253uo  
 APPLICANT: NAKAZATO, Hiroshi  
 APPLICANT: MIURA, Kenju  
 APPLICANT: ISHIDA, No. 5874253uhiro  
 APPLICANT: KURIHARA, Tatsuya  
 APPLICANT: YAMAGUCHI, Kozo  
 APPLICANT: YAMAGUCHI, No. 5874253omi  
 TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
 NUMBER OF SEQUENCES: 34  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Burns, Doane, Swecker & Mathis  
 STREET: George Mason Bldg., Washington & Prince Sts.  
 City: Alexandria  
 STATE: Virginia  
 COUNTRY: United States  
 ZIP: 22313-1404  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentia Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/474,661  
 FILING DATE: 07-JUN-1995  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/091,028

FILING DATE: 14-JUL-1993  
 APPLICATION NUMBER: JP 4-212305  
 FILING DATE: 17-JUL-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: JP 5-067339  
 FILING DATE: 04-MAR-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: REA, TERESA STANEK  
 REGISTRATION NUMBER: 30,427  
 REFERENCE/DOCKET NUMBER: 001560-204  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 836-6620  
 TELEFAX: (703) 836-6620  
 INFORMATION FOR SEQ. ID NO: 34:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 380 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-474-661-34

Query Match 100.0%; Score 1959; DB 2; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 1e-178;  
 Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASLAANAEFCNLFREMDNNGNNGVFFSSLSLPAALALVRLGAQDDSLQIDKLHY 60  
 DB 1 MASLAANAEFCNLFREMDNNGNNGVFFSSLSLPAALALVRLGAQDDSLQIDKLHY 60  
 QY 61 NTASGYGNSNSGSGLOSOLKRVFSDINASHKDYDLSYNGLEPAEKYGFHKYIECAEK 120  
 DB 61 NTASGYGNSNSGSGLOSOLKRVFSDINASHKDYDLSYNGLEPAEKYGFHKYIECAEK 120  
 QY 121 LYDAKVERVDFTNHLIEDTRRNINKWVENETHGKIKNVIGSGISSAVMLVNAVYFEGK 180  
 DB 121 LYDAKVERVDFTNHLIEDTRRNINKWVENETHGKIKNVIGSGISSAVMLVNAVYFEGK 180  
 QY 181 WQSAFTKSETINCHFKSPKCSGKAVAMHQBKFNLSYIEDPSMKILIELRYNGINMYVL 240  
 DB 181 WQSAFTKSETINCHFKSPKCSGKAVAMHQBKFNLSYIEDPSMKILIELRYNGINMYVL 240  
 QY 241 LPENDLSEIENKLTFOIEMETNPRRMTSKYVEVFFPOFKIEKNYEMKOYLRLGLKIDIF 300  
 DB 241 LPENDLSEIENKLTFOIEMETNPRRMTSKYVEVFFPOFKIEKNYEMKOYLRLGLKIDIF 300  
 QY 301 DESKADLSGIASGRLYISMMHKSYTEVEEGTEATAATGSNIVEKOLPOSTLFRADHP 360  
 DB 301 DESKADLSGIASGRLYISMMHKSYTEVEEGTEATAATGSNIVEKOLPOSTLFRADHP 360  
 QY 361 FLFVIRKDDIILFSGKVS 380  
 DB 361 FLFVIRKDDIILFSGKVS 380

## RESULT 3

US-08-611-977-34  
 Sequence 34, Application US/08611977  
 Patent No. 5972886  
 GENERAL INFORMATION:  
 APPLICANT: TSUJIMOTO, Masafumi  
 APPLICANT: IWASA, Fuyuki  
 APPLICANT: TSUBUOKA, No. 5972886uo  
 APPLICANT: NAKAZATO, Hiroshi  
 APPLICANT: MIURA, Kenju  
 APPLICANT: KURIHARA, Tatsuya  
 APPLICANT: ISHIDA, No. 5972886uhiro  
 APPLICANT: KURIHARA, Tatsuya  
 APPLICANT: YAMAGUCHI, Kozo  
 APPLICANT: YAMAGUCHI, No. 5972886omi  
 TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
 NUMBER OF SEQUENCES: 34  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Burns, Doane, Swecker & Mathis  
 STREET: P.O. Box 1404

CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/611,977  
FILING DATE: 06-MAR-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-204  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-611-977-34

Query Match 100.0%; Score 1959; DB 2; Length 380;  
Best Local Similarity 100.0%; Pred. No. 1e-178;  
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASLAANAEEFCNLFREMDNOCNGNVFFSSLSLFAALALVRLGAODDSLQIDKLHHV 60  
DB 1 MASLAANAEEFCNLFREMDNOCNGNVFFSSLSLFAALALVRLGAODDSLQIDKLHHV 60  
QY 61 NTASGYGNSSNSQSGLOSLKRVPSDINASHKDYDLSIVNGLPAEKYGFHKKDYIECAEK 120  
DB 61 NTASGYGNSSNSQSGLOSLKRVPSDINASHKDYDLSIVNGLPAEKYGFHKKDYIECAEK 120  
QY 121 LYDAKVERVDFTNHLIEDTRRNINKWVENETHGKIKNVIGGGISSAVMYLVNMYVFKGK 180  
DB 121 LYDAKVERVDFTNHLIEDTRRNINKWVENETHGKIKNVIGGGISSAVMYLVNMYVFKGK 180  
QY 181 MOSAFTSETINCHFKSPKCSGKAVAMHOERKENLSVIDPSSKILIELRYNGGINMYVL 240  
DB 181 MOSAFTSETINCHFKSPKCSGKAVAMHOERKENLSVIDPSSKILIELRYNGGINMYVL 240  
QY 241 LPENDLSEINLKLTFQNLMEWTNPRMTSKYVEVFPOFKIEKNYEMKQYLRALGLDIF 300  
DB 241 LPENDLSEINLKLTFQNLMEWTNPRMTSKYVEVFPOFKIEKNYEMKQYLRALGLDIF 300  
QY 301 DESKADISGIASGRGLYSRMHKSYLEVTEEGTEATATGNSNVEKOLPOSTLFRADHP 360  
DB 301 DESKADISGIASGRGLYSRMHKSYLEVTEEGTEATATGNSNVEKOLPOSTLFRADHP 360  
QY 361 FLFVIRKDDIILFSGKVS CP 380  
DB 361 FLFVIRKDDIILFSGKVS CP 380

RESULT 4  
US-08-568-147B-2  
; Sequence 2, Application US/08568147B

Patent No. 5783422  
GENERAL INFORMATION:  
APPLICANT: Sumitani, Yoshinori  
APPLICANT: Kato, Hiroshi  
APPLICANT: Sekiguchi, Kiyoshi  
APPLICANT: Takeda, Katsumichi  
TITLE OF INVENTION: DNA FRAGMENT CODING FOR SQUAMOUS CELL  
TITLE OF INVENTION: CARCINOMA-ASSOCIATED ANTIGEN  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Scully, Scott, Murphy & Presser  
STREET: 400 Garden City Plaza  
CITY: Garden City  
STATE: New York  
COUNTRY: USA  
ZIP: 11530  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/568,147B  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 800,952  
FILING DATE: 02-DEC-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Digilio, Frank S.  
REGISTRATION NUMBER: 31,346  
REFERENCE/DOCKET NUMBER: 8425  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 516-742-4343  
TELEFAX: 516-742-4366  
TELEX: 230 901 SANS UR  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 390 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-568-147B-2

Query Match 38.4%; Score 752; DB 1; Length 390;  
Best Local Similarity 41.5%; Pred. No. 1.6e-63;  
Matches 165; Conservative 81; Mismatches 126; Indels 26; Gaps 10;

QY 1 MASLAANAEEFCNLFREMDNOCNGNVFFSSLSLFAALALVRLGAODDSLQIDKLHHV 60  
DB 1 MNSLSEANTKMFPLFOQFRKSKEN-NIFSPISITSLGVLNVLGAKDNTAQQIKKYLHF 59  
QY 61 -----NTASGYGNSSNSQSG-LQSLKRVPSDINASHKDYDLSIVNGLPAEKYGFHKKDY 114  
DB 60 DQVTEHTTGKATYHVRSGGVHHQFQKILTEFNKSTDAVELKLANLKFGEKTYLFLQEY 119  
QY 115 IECAEKLYDAKVERVDFTNHLIEDTRRNINKWVENETHGKIKNVIGGGISSAVMYLVN 174  
DB 120 LDAIKKFFQYQSVESVDANAPPEESKRKINSVVESSQYNEKINLIPBGIGSNTTLVYNA 179  
QY 175 VYFKGWSAFTKSETINCHFKSPKCSGKAVAMHOERKENLSVIDPSSKILIELRYNG- 233  
DB 180 IYFKGWEKFKNKEDITEKFTWPKNTYKSIOMHROTTSFPAILEDVQAKVLEIPKKG 239  
QY 234 GINMYVLPE--NDLSEINLKLTFQNLMEWTNPRMTSKYVEVFPOFKIEKNYEMKQYL 291  
DB 240 DLSMIVLLPNTIDGLQLEKLTAEKLMWTSLSQNMRETRVDLHPKRYEESYDLKDTL 299  
QY 292 RALGLKIDFDESKADISGIASGRGLYSRMHKSYLEVTEEGTATATAT-----GSNIVE 346  
DB 300 RTMGMDVIFN-GDADISGIMTGSRLGLVSLGAKAFVEYTEEGCAAAATAYVGGSS-- 355  
QY 347 KQLPOST--LFRADHPFLFVIR--KDDIILFSGKVS CP 380





CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/121.714  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/938,823  
FILING DATE: 09/01/92  
APPLICATION NUMBER: 07/844,296  
FILING DATE: 02/28/92  
APPLICATION NUMBER: 07/662,216  
FILING DATE: 02/28/91  
ATTORNEY/AGENT INFORMATION:  
NAME: Fraser, Janis K.  
REGISTRATION NUMBER: 34,819  
REFERENCE/DOCKET NUMBER: 00530/072001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 375  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-08-121-714-8

Query Match 37.3%; Score 730.5; DB 1; Length 375;  
Best Local Similarity 37.9%; Pred. No. 1.7e-61;  
Matches 146; Conservative 85; Mismatches 135; Indels 19; Gaps 4;

QY 1 MASLAANAEEFCFLPREMDNNGNNGVFFSSLSLFAALALVRLGAGDDSLSDIKLHV 60  
DB 1 MEOLSTANTHFAVLDFRALNEDPTGNIFFISPLISSSLAMIFLGTGNTAAYSKALYF 60  
QY 61 NTASGYGNSNSQSGLSQKRVFSDINASHKDYDLISVNGLFAEKYVGFHKDYIECAEK 120  
DB 61 DTV-----EDHSRFQSLNADINKPCAPYILKLANLYGEKTYNFLADFLASQK 110  
QY 121 LYDAKVERDFTNHLNEDTRRNINKVVENETGKIKNYIGEGCISSAAMVLVNAVYERK 180  
DB 111 MYGAELIASVDFQOAPEDARKKEINEMVKGOTGKIPDLVYKQMVNMKTLVNAVYERK 170  
QY 181 WQAFKSETINCHFKSPKSGKAVAMHOBKRNLSYIEDPSMKILELRVNG-GINNVY 239  
DB 171 WQEFKMATRDAPFLNKKDTKTKVKKMYOKKKPPYNIEDLCKRVLELPQGEKLSMTI 230  
QY 240 LLPEN-----DLSEIENKLTFOULMEWTNPRMTSKYVEVFFPOFKIEKNYEMKOYLRA 293  
DB 231 LLPDIEDESTGLEKIEKQLFLKLRWTKRENLYLAENVHLPFRKLEESYDLTSHLAR 290  
QY 294 LGLKDIDESKADLSGIASGRLYISRMHKSYLEVEEGTEATPAAGSNIVEKQLPOST 353  
DB 291 LGVODLFNRGRKADLSGMSGARDLFVSKIIHKSFPVDLNEEGTEAATAAGTITMLAMPBE 350  
QY 354 LFRADHPELFVIRKDDI--ILFSCK 376  
DB 351 NFADHPETFFIRHNPANILFLGR 375

RESULT 8  
US-08-477-108A-8  
Sequence 8, Application US/08477108A  
Patent No. 5801001  
GENERAL INFORMATION:  
APPLICANT: Sager, Ruth  
APPLICANT: 'Zou, Zhigang  
APPLICANT: Anisowicz, Anthony  
TITLE OF INVENTION: MASPIN, A NOVEL SERPIN WITH  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.

STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM PS/2 Model 502 or 55SX  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: WordPerfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/477.108A  
FILING DATE: June 7, 1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/121.714  
FILING DATE: 09/01/93  
APPLICATION NUMBER: 07/938,823  
FILING DATE: 09/01/92  
APPLICATION NUMBER: 07/844,296  
FILING DATE: 02/28/92  
APPLICATION NUMBER: 07/662,216  
FILING DATE: 02/28/91  
ATTORNEY/AGENT INFORMATION:  
NAME: Fraser, Janis K.  
REGISTRATION NUMBER: 34,819  
REFERENCE/DOCKET NUMBER: 06570/002002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 375  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-08-477-108A-8

Query Match 37.3%; Score 730.5; DB 1; Length 375;  
Best Local Similarity 37.9%; Pred. No. 1.7e-61;  
Matches 146; Conservative 85; Mismatches 135; Indels 19; Gaps 4;

QY 1 MASLAANAEEFCFLPREMDNNGNNGVFFSSLSLFAALALVRLGAGDDSLSDIKLHV 60  
DB 1 MEOLSTANTHFAVLDFRALNEDPTGNIFFISPLISSSLAMIFLGTGNTAAYSKALYF 60  
QY 61 NTASGYGNSNSQSGLSQKRVFSDINASHKDYDLISVNGLFAEKYVGFHKDYIECAEK 120  
DB 61 DTV-----EDHSRFQSLNADINKPCAPYILKLANLYGEKTYNFLADFLASQK 110  
QY 121 LYDAKVERDFTNHLNEDTRRNINKVVENETGKIKNYIGEGCISSAAMVLVNAVYERK 180  
DB 111 MYGAELIASVDFQOAPEDARKKEINEMVKGOTGKIPDLVYKQMVNMKTLVNAVYERK 170  
QY 181 WQAFKSETINCHFKSPKSGKAVAMHOBKRNLSYIEDPSMKILELRVNG-GINNVY 239  
DB 171 WQEFKMATRDAPFLNKKDTKTKVKKMYOKKKPPYNIEDLCKRVLELPQGEKLSMTI 230  
QY 240 LLPEN-----DLSEIENKLTFOULMEWTNPRMTSKYVEVFFPOFKIEKNYEMKOYLRA 293  
DB 231 LLPDIEDESTGLEKIEKQLFLKLRWTKRENLYLAENVHLPFRKLEESYDLTSHLAR 290  
QY 294 LGLKDIDESKADLSGIASGRLYISRMHKSYLEVEEGTEATPAAGSNIVEKQLPOST 353  
DB 291 LGVODLFNRGRKADLSGMSGARDLFVSKIIHKSFPVDLNEEGTEAATAAGTITMLAMPBE 350  
QY 354 LFRADHPELFVIRKDDI--ILFSCK 376  
DB 351 NFADHPETFFIRHNPANILFLGR 375

RESULT 9



Db 171 WOEFMKEATRADAPFRLNKDKTKYKMMYQKKKFPYNIEDLKCRVLELPYQKGLSMII 230  
Qy 240 LLPEN-----DLSEIENKLTFFONLMEWTNPRMTSKYVEVFFPOFKIEKNYEMKOYRA 293  
Db 231 LLPDDIEDESGLEKIEQLTLEKLTREWTKPENLYLAENVHLPEFKLEESYDLTSLHAR 290  
Qy 294 LGLADIPESKADLSGASGRLYISRMHKSXYIEVTEEGTEATGATGSNIVEKOLPOST 353  
Db 291 LGVODLFNRKADLSGMSGARDLVYSKIHKSPVDLNEEGTEAATACTIMLANLMPPE 350  
Qy 354 LFRADHPLFVIRKDDI--ILFSGK 376  
Db 351 NFNADHPLEFIRHNPSANILFLGR 375

RESULT 11  
US-08-121-714-4  
; Sequence 4, Application US/08121714  
; Patent No. 5470970  
; GENERAL INFORMATION:  
; APPLICANT: Sager, Ruth  
; TITLE OF INVENTION: MASPIN, A NOVEL SERPIN WITH  
; TITLE OF INVENTION: TUMOR SUPPRESSING ACTIVITY  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; COMPUTER: IBM PS/2 Model 50Z or 55SX  
; OPERATING SYSTEM: MS-DOS (Version 5.0)  
; SOFTWARE: WordPerfect (Version 5.1)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/121,714  
; FILING DATE:  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/938, 823  
; FILING DATE: 09/01/92  
; APPLICATION NUMBER: 07/844, 296  
; FILING DATE: 02/28/92  
; APPLICATION NUMBER: 07/662, 216  
; FILING DATE: 02/28/91  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fraser, Janis K.  
; REGISTRATION NUMBER: 34, 819  
; REFERENCE/DOCKET NUMBER: 00530/072001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 542-5070  
; TELEFAX: (617) 542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 379  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
US-08-121-714-4

Query Match 36.9%; Score 722.5; DB 1; Length 379;  
Best Local Similarity 38.6%; Pred. No. 1e-60;  
Matches 150; Conservative 74; Mismatches 146; Indels 19; Gaps 4;

Qy 1 MASLAANAEEFCNLFREMDNOCNGNVEFFSLSFALALVRLGAODSDLSQIDKLHAY 60  
Db 1 MEOLSSANTRALDLFLALSENPNAGNIFISPFSSISAMAVNGLTGRNTAQLSKTCHF 60  
Qy 61 NTASGYGNSNSQSLQSLKRVFSDINASHKDYDLSTVNGLFAEKVYGFHHDYIECAEK 120

Db 61 NTV-----EEVHSRFSQSLNADINKRGASYLILKLANLRYGKTYNPLPEFLVSTOK 110  
Qy 121 LYDAKVERVDFTNHLIEDTRRNINKVYENETGKINAVYEGGSISSAMVLYNAVYFGK 180  
Db 111 TYGADLAVSDVQAHASEDARKTINOVVKQOTGKIPELLASQWDMVTLYVNAVYFGN 170  
Qy 181 WSAFTKETETINCHFSKPCSGKAVAMHOBKRFNLSTYEDPSMKILELRNG-GINNVY 239  
Db 171 WKDFMKEATINAPFRLNKDKTKYKMMYQKKKFPYNIEDLKCRVLELPYQKGLSMII 230  
Qy 240 LLPEN-----DLSEIENKLTFFONLMEWTNPRMTSKYVEVFFPOFKIEKNYEMKOYRA 293  
Db 231 LLPDDIEDESGLEKIEQLTLEKLTREWTKPENLDFIENVSLPREFKLEESYTLNSDLAR 290  
Qy 294 LGLADIPESKADLSGASGRLYISRMHKSXYIEVTEEGTEATGATGSNIVEKOLPOST 353  
Db 291 LGVODLFNRKADLSGMSGARDLVYSKIHKSPVDLNEEGTEAATACTIMLANLMPPE 350  
Qy 354 LFRADHPLFVIRKDDI--ILFSGK 380  
Db 351 NFNADHPLEFIRHNPSANILFLGRFSSP 379

RESULT 12  
US-08-477-108A-4  
; Sequence 4, Application US/08477108A  
; Patent No. 5801001  
; GENERAL INFORMATION:  
; APPLICANT: Sager, Ruth  
; APPLICANT: Zou, Zhigiang  
; APPLICANT: Anisowicz, Anthony  
; TITLE OF INVENTION: MASPIN, A NOVEL SERPIN WITH  
; TITLE OF INVENTION: TUMOR SUPPRESSING ACTIVITY  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; COMPUTER: IBM PS/2 Model 50Z or 55SX  
; OPERATING SYSTEM: MS-DOS (Version 5.0)  
; SOFTWARE: WordPerfect (Version 5.1)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/477, 108A  
; FILING DATE: June 7, 1995  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/121, 714  
; FILING DATE: 09/01/93  
; APPLICATION NUMBER: 07/938, 823  
; FILING DATE: 09/01/92  
; APPLICATION NUMBER: 07/844, 296  
; FILING DATE: 02/28/92  
; APPLICATION NUMBER: 07/662, 216  
; FILING DATE: 02/28/91  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fraser, Janis K.  
; REGISTRATION NUMBER: 34, 819  
; REFERENCE/DOCKET NUMBER: 06570/002002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 542-5070  
; TELEFAX: (617) 542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 379  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear

US-08-477-108A-4

Query Match	36.9%	Score 722.5;	DB 1;	length 379;
Best Local Similarity	38.6%	Pred. No. 1e-60;		
Matches 150;	Conservative 74;	Mismatches 146;	Indels 19;	Gaps 4

RESULT 13  
 US-08/477-112-4  
 : Sequence 4, Application US/08/477112  
 : Patent No. 5905023  
 :  
 GENERAL INFORMATION:  
 APPLICANT: Sager, Ruth  
 TITLE OF INVENTION: MASPIN, A NOVEL SERPIN WITH  
 TITLE OF INVENTION: TUMOR SUPPRESSING ACTIVITY  
 NUMBER OF SEQUENCES: 8  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Fish & Richardson P.C.  
 STREET: 225 Franklin Street  
 CITY: Boston  
 STATE: Massachusetts  
 COUNTRY: U.S.A.  
 ZIP: 02110-2804  
 :  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
 COMPUTER: IBM PS/2 Model 50z or 55SX  
 OPERATING SYSTEM: MS-DOS (Version 5.0)  
 SOFTWARE: WordPerfect (Version 5.1)  
 :  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/477,112  
 FILING DATE:  
 CLASSIFICATION: 536  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/121,714  
 FILING DATE: 09/01/93  
 APPLICATION NUMBER: 07/938,823  
 FILING DATE: 09/01/92  
 APPLICATION NUMBER: 07/844,296  
 FILING DATE: 02/28/92  
 APPLICATION NUMBER: 07/662,216  
 FILING DATE: 02/28/91  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Fraser, Janis K.  
 REGISTRATION NUMBER: 34,819  
 REFERENCE/DOCKET NUMBER: 06570/002003

TELECOMMUNICATION INFORMATION  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
SEQUENCE CHARACTERISTICS:

US-08-477-112

Query Match	36.9%	Score 722.5;	DB 2;	Length 379;
Best Local Similarity	38.6%;	Pred. No. 1e-60;		
Matches 150;	Conservative 74;	Mismatches 146;	Indels 19;	Gaps 4;

RESULT 14  
PCT-US93-08322-4  
Sequence 4, Application PC/TUS9308322  
GENERAL INFORMATION:  
APPLICANT: Sager, Ruth  
TITLE OF INVENTION: MASPIN, A NOVEL SERPIN WITH TUMOR SUPPRESSING ACTIVITY  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM PS/2 Model 502 or 555X  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: WordPerfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/08322  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/938,823  
FILING DATE: 09/01/92  
APPLICATION NUMBER: 07/844,296  
FILING DATE: 02/28/92



**THIS PAGE BLANK (USPTO)**

Fri Jul 11 15:10:51 2003

us-10-091-442-34.rapb

Page 1

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:59:17 ; Search time 111.393 Seconds  
(without alignments)  
397.148 Million cell updates/sec

Title: US-10-091-442-34  
Perfect score: 1959  
Sequence: 1 MASLAAMNAFCNLFREMD.....FLFYIRKDDIILFSGKYSVP 380

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 445758 seqs, 11641973 residues  
Total number of hits satisfying chosen parameters: 445758

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published.Applications\_AA:\*

- 1: /cgn2\_6/ptodata/1/pubppaa/US07\_NEM\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubppaa/PCRT\_NEM\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubppaa/US06\_NEM\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubppaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubppaa/PCRTUS\_PUBCOMB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubppaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubppaa/US08\_NEM\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubppaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubppaa/US09\_NEM\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubppaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubppaa/US10\_NEM\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubppaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubppaa/US60\_NEM\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubppaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1959	100.0	380	9 US-10-091-442-34	Sequence 34, Appl
2	1959	100.0	380	9 US-09-140-719-34	Sequence 34, Appl
3	822	42.0	392	9 US-10-198-070-62	Sequence 62, Appl
4	820	41.9	392	9 US-10-198-070-67	Sequence 67, Appl
5	819	41.8	392	9 US-10-198-070-70	Sequence 70, Appl
6	796.5	40.7	379	9 US-10-012-896-946	Sequence 946, App
7	796.5	40.7	379	9 US-09-895-814-946	Sequence 946, App
8	771.5	39.4	395	9 US-09-823-187-2	Sequence 2, Appl
9	752	38.4	390	9 US-09-823-187-29	Sequence 29, Appl
10	752	38.4	390	9 US-09-823-187-30	Sequence 29, Appl
11	748	38.2	390	9 US-09-823-187-27	Sequence 27, Appl
12	743	37.9	390	9 US-09-823-187-28	Sequence 27, Appl
13	740	37.8	390	7 US-08-731-566-2	Sequence 2, Appl
14	738.5	37.7	617	9 US-10-012-896-947	Sequence 947, App
15	738.5	37.7	617	9 US-08-895-814-947	Sequence 947, App
16	732.5	36.9	437	10 US-09-925-301-1200	Sequence 1200, App
17	715.5	36.5	391	9 US-10-157-031-355	Sequence 355, App
18	715.5	36.5	391	9 US-09-466-396A-110	Sequence 110, App
19	715.5	36.5	391	10 US-09-735-705-110	Sequence 110, App

20	715.5	36.5	391	10 US-09-850-716A-110	Sequence 110, App
21	715.5	36.5	391	10 US-09-897-778-110	Sequence 110, App
22	715	36.5	400	9 US-09-466-396A-112	Sequence 112, App
23	715	36.5	400	10 US-09-735-705-112	Sequence 112, App
24	715	36.5	400	10 US-09-850-716A-112	Sequence 112, App
25	715	36.5	400	10 US-09-897-778-112	Sequence 112, App
26	672.5	34.3	340	10 US-09-764-898-169	Sequence 169, App
27	664.5	33.9	415	9 US-10-017-128-4	Sequence 4, Appl
28	664.5	33.9	423	9 US-10-106-658-4538	Sequence 4538, App
29	662.5	33.8	415	10 US-09-902-684-4	Sequence 4, Appl
30	661	33.7	374	9 US-09-974-298-95	Sequence 95, Appl
31	654	33.4	385	9 US-09-847-208-101	Sequence 101, App
32	652	33.3	409	9 US-10-267-311-55	Sequence 55, Appl
33	652	33.3	948	9 US-10-267-311-21	Sequence 21, Appl
34	642	32.8	459	10 US-09-925-300-1440	Sequence 1440, App
35	608	31.0	315	7 US-08-731-566-4	Sequence 4, Appl
36	600.5	30.7	377	9 US-09-823-187-31	Sequence 31, Appl
37	592	30.2	465	9 US-09-987-021-6	Sequence 6, Appl
38	592	30.2	465	10 US-09-957-485-6	Sequence 6, Appl
39	566	28.9	464	10 US-09-414-834-1	Sequence 21, Appl
40	544	27.8	377	9 US-10-165-605A-27	Sequence 27, Appl
41	544	27.8	377	10 US-09-910-430-27	Sequence 27, Appl
42	498.5	25.4	410	9 US-09-987-021-2	Sequence 2, Appl
43	498.5	25.4	410	10 US-09-957-485-2	Sequence 2, Appl
44	477.5	24.4	217	10 US-09-729-674-134	Sequence 134, App
45	476.5	24.3	402	9 US-10-135-629-4	Sequence 4, Appl

#### ALIGNMENTS

RESULT 1  
US-10-091-442-34  
Sequence 34, Application US/10091442  
Patent No. US20020164711A1  
GENERAL INFORMATION:  
APPLICANT: TSUTSUMOTO, Masafumi  
IWASA, Fuyuki  
TSURUOKA, No. US20020164711A1uo  
NAKAZATO, Hiroshi  
MURA, Kenju  
ISHIDA, No. US20020164711A1uhfro  
KURIHARA, Tatsuya  
YAMACHI, Kozo  
YAMAGUCHI, No. US20020164711A1omh  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22113-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/091,442  
FILING DATE: 07-Mar-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-Aug-1998  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-Jun-1995  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-Jul-1993  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-Jul-1992  
APPLICATION NUMBER: JP 6-067339

FILING DATE: 04-MAR-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: McGowan, Malcolm K.  
 REGISTRATION NUMBER: 39,300  
 REFERENCE/DOCKET NUMBER: 001560-247  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 836-6620  
 TELEFAX: (703) 836-2021  
 INFORMATION FOR SEQ ID NO: 34:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 380 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 SEQUENCE DESCRIPTION: SEQ ID NO: 34:  
 US-10-091-442-34

Query Match 100.0%; Score 1959; DB 9; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 6,8e-161;  
 Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASLAANAEEFCNLFREMDNDGNGNVEFSSLSLFAALALVRLGAODDSLQIDKLLHY 60  
 1 MASLAANAEEFCNLFREMDNDGNGNVEFSSLSLFAALALVRLGAODDSLQIDKLLHY 60  
 DB 1 MASLAANAEEFCNLFREMDNDGNGNVEFSSLSLFAALALVRLGAODDSLQIDKLLHY 60  
 QY 61 NTASGYGNSNSGSGLOSLKRVFSDINASHKDYDSIYNGLFAEKYVGFHKDYIECAEK 120  
 61 NTASGYGNSNSGSGLOSLKRVFSDINASHKDYDSIYNGLFAEKYVGFHKDYIECAEK 120  
 DB 61 NTASGYGNSNSGSGLOSLKRVFSDINASHKDYDSIYNGLFAEKYVGFHKDYIECAEK 120  
 QY 121 LYDAKERYVDFTNHLDTNRINKWVENETHGKIKNVIGEGISSAVMVLVNAVYFKGK 180  
 121 LYDAKERYVDFTNHLDTNRINKWVENETHGKIKNVIGEGISSAVMVLVNAVYFKGK 180  
 DB 121 LYDAKERYVDFTNHLDTNRINKWVENETHGKIKNVIGEGISSAVMVLVNAVYFKGK 180  
 QY 181 WQSAFTKSETINCHFSKPCSGKAVAMMOERKFNLSYIEDPSMKLLELRNGGINMYVL 240  
 181 WQSAFTKSETINCHFSKPCSGKAVAMMOERKFNLSYIEDPSMKLLELRNGGINMYVL 240  
 DB 181 WQSAFTKSETINCHFSKPCSGKAVAMMOERKFNLSYIEDPSMKLLELRNGGINMYVL 240  
 QY 241 LPENDLSEIENKLTFFONLMEWTPRRMTSKYVEVFPOPKIEKNYEMKOYLALGLKIDF 300  
 241 LPENDLSEIENKLTFFONLMEWTPRRMTSKYVEVFPOPKIEKNYEMKOYLALGLKIDF 300  
 DB 241 LPENDLSEIENKLTFFONLMEWTPRRMTSKYVEVFPOPKIEKNYEMKOYLALGLKIDF 300  
 QY 301 DESKADLSGIASGRLYISRMHMKSYIEVTEEGTEATAAGSNIVEKQLPOSTLFRADHP 360  
 301 DESKADLSGIASGRLYISRMHMKSYIEVTEEGTEATAAGSNIVEKQLPOSTLFRADHP 360  
 DB 301 DESKADLSGIASGRLYISRMHMKSYIEVTEEGTEATAAGSNIVEKQLPOSTLFRADHP 360  
 QY 361 FLFVIRKDDIILFSGKVS CP 380  
 361 FLFVIRKDDIILFSGKVS CP 380  
 DB 361 FLFVIRKDDIILFSGKVS CP 380

RESULT 2  
 US-09-140-719-34  
 Sequence 34, Application US/09140719  
 Patent No. US20010026931A1  
 GENERAL INFORMATION:  
 APPLICANT: TSUJIMOTO, Masafumi  
 APPLICANT: IWASA, Fuyuki  
 APPLICANT: TSUROOKA, No. US20010026931A1uo  
 APPLICANT: NAKAZATO, Hiroshi  
 APPLICANT: MIURA, Kenju  
 APPLICANT: ISHIDA, No. US20010026931A1uhtro  
 APPLICANT: KURIHARA, Tatsuya  
 APPLICANT: YAMAGUCHI, Kozo  
 APPLICANT: YAMAGUCHI, No. US20010026931A1om1  
 TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
 NUMBER OF SEQUENCES: 34  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Burns, Doane, Swecker & Mathis  
 STREET: P.O. Box 1404  
 CITY: Alexandria  
 STATE: Virginia  
 COUNTRY: United States  
 ZIP: 22313-1404

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/140,719  
 FILING DATE: 08-AUG-1998  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/474,661  
 FILING DATE: 07-JUN-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/091,028  
 FILING DATE: 14-JUL-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: JP 4-212305  
 FILING DATE: 17-JUL-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: JP 6-067339  
 FILING DATE: 04-MAR-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: McGowan, Malcolm K.  
 REGISTRATION NUMBER: 39,300  
 REFERENCE/DOCKET NUMBER: 001560-247  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 836-6620  
 TELEFAX: (703) 836-2021  
 INFORMATION FOR SEQ ID NO: 34:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 380 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-09-140-719-34

Query Match 100.0%; Score 1959; DB 10; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 6,8e-161;  
 Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASLAANAEEFCNLFREMDNDGNGNVEFSSLSLFAALALVRLGAODDSLQIDKLLHY 60  
 1 MASLAANAEEFCNLFREMDNDGNGNVEFSSLSLFAALALVRLGAODDSLQIDKLLHY 60  
 DB 1 MASLAANAEEFCNLFREMDNDGNGNVEFSSLSLFAALALVRLGAODDSLQIDKLLHY 60  
 QY 61 NTASGYGNSNSGSGLOSLKRVFSDINASHKDYDSIYNGLFAEKYVGFHKDYIECAEK 120  
 61 NTASGYGNSNSGSGLOSLKRVFSDINASHKDYDSIYNGLFAEKYVGFHKDYIECAEK 120  
 DB 61 NTASGYGNSNSGSGLOSLKRVFSDINASHKDYDSIYNGLFAEKYVGFHKDYIECAEK 120  
 QY 121 LYDAKERYVDFTNHLDTNRINKWVENETHGKIKNVIGEGISSAVMVLVNAVYFKGK 180  
 121 LYDAKERYVDFTNHLDTNRINKWVENETHGKIKNVIGEGISSAVMVLVNAVYFKGK 180  
 DB 121 LYDAKERYVDFTNHLDTNRINKWVENETHGKIKNVIGEGISSAVMVLVNAVYFKGK 180  
 QY 181 WQSAFTKSETINCHFSKPCSGKAVAMMOERKFNLSYIEDPSMKLLELRNGGINMYVL 240  
 181 WQSAFTKSETINCHFSKPCSGKAVAMMOERKFNLSYIEDPSMKLLELRNGGINMYVL 240  
 DB 181 WQSAFTKSETINCHFSKPCSGKAVAMMOERKFNLSYIEDPSMKLLELRNGGINMYVL 240  
 QY 241 LPENDLSEIENKLTFFONLMEWTPRRMTSKYVEVFPOPKIEKNYEMKOYLALGLKIDF 300  
 241 LPENDLSEIENKLTFFONLMEWTPRRMTSKYVEVFPOPKIEKNYEMKOYLALGLKIDF 300  
 DB 241 LPENDLSEIENKLTFFONLMEWTPRRMTSKYVEVFPOPKIEKNYEMKOYLALGLKIDF 300  
 QY 301 DESKADLSGIASGRLYISRMHMKSYIEVTEEGTEATAAGSNIVEKQLPOSTLFRADHP 360  
 301 DESKADLSGIASGRLYISRMHMKSYIEVTEEGTEATAAGSNIVEKQLPOSTLFRADHP 360  
 DB 301 DESKADLSGIASGRLYISRMHMKSYIEVTEEGTEATAAGSNIVEKQLPOSTLFRADHP 360  
 QY 361 FLFVIRKDDIILFSGKVS CP 380  
 361 FLFVIRKDDIILFSGKVS CP 380  
 DB 361 FLFVIRKDDIILFSGKVS CP 380

RESULT 3  
 US-10-198-070-62  
 Sequence 62, Application US/10198070  
 Publication No. US20030109437A1



GENERAL INFORMATION:  
APPLICANT: AVERBACK, PAUL  
APPLICANT: GEMMELL, JACK  
TITLE OF INVENTION: PEPTIDES EFFECTIVE IN THE TREATMENT OF TUMORS AND OTHER  
TITLE OF INVENTION: CONDITIONS REQUIRING THE REMOVAL OR DESTRUCTION OF  
TITLE OF INVENTION: CELLS  
FILE REFERENCE: 59003.000008  
CURRENT APPLICATION NUMBER: US/10/198, 070  
CURRENT FILING DATE: 2002-07-19  
PRIOR APPLICATION NUMBER: 60/306,161  
PRIOR FILING DATE: 2001-07-19  
PRIOR APPLICATION NUMBER: 60/306,150  
PRIOR FILING DATE: 2001-07-19  
PRIOR APPLICATION NUMBER: 60/331,477  
PRIOR FILING DATE: 2001-11-16  
NUMBER OF SEQ ID NOS: 125  
SOFTWARE: Patentln Ver. 2.1  
SEQ ID NO 62  
LENGTH: 392  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-198-070-62

Query Match 42.0%; Score 822; DB 9; Length 392;  
Best Local Similarity 43.1%; Pred. No. 8.5e-63;  
Matches 169; Conservative 80; Mismatches 131; Indels 12; Gaps 6;

OY 1 MASIAANAEFCNLFREMDNONGVNFSSLSLFAALALVRLGADDSISQIDKLLHV 60  
DB 1 MGSISTANVEFCIDVFEKLSNNIGDINFSSLSLALSLVLLGARGETAOLEKVLHF 60  
OY 61 -----NTASGYGNSN-SOSG-LOSOLKRVESDINASHKDYLSIVGLFAEKYGGPHKD 113  
DB 61 SHYVDSLKPFKDSQKSGRHSFGEVSQJNODSCTLSIANRLVGTKTMATHQ 120  
OY 114 YIECAERLYDAKERVDFTHLEDTRRNKKNVENETHGRIKINVGIGISSAVMYLVN 173  
DB 121 YLSCSEKMYQARLQTVDFEQSTETRTKINAMWENKNGKVANLFKSTIDPSSVAVLVN 180  
OY 174 AVYFKGKQSAFTSETINCHFKSPKSGKAVAMHDEKFNLSVIDPSSKILELY-N 232  
DB 181 TTYFGGROKMFQYREVKSPFQISEGNVYEMMYOIGTFKLAVEREPOQVLELYPVN 240  
OY 233 GGIMNYVLLPEN--DLSEIENKLFONLMEWTNPRRTSKYEVFFPOFKIEKNYEMKQY 290  
DB 241 NKLSMITLLPVGIANLQIEKQNSGFHEHTSSNMREVEVHLRFKILEITYELNSL 300  
OY 291 LRALGLDIFDESKADLSGASGRLYISRMHKSLEYTEEGTEATAGSNIVKQLP 350  
DB 301 LKPLGVTDLFNOVKADLSGMSPTKGLYLSKAIRKSYLDVSEGEAAATGDSIAVKSPL 360  
OY 351 OSTLFRAHDFPLFYIR--KDDILFSGKVSQP 380  
DB 361 MRAQFKANHPFLFFIRHTHTNTLLFCGLASP 392

RESULT 4  
US-10-198-070-67  
Sequence 67, Application US/10198070  
Publication No. US20030109437A1  
GENERAL INFORMATION:  
APPLICANT: AVERBACK, PAUL  
APPLICANT: GEMMELL, JACK  
TITLE OF INVENTION: PEPTIDES EFFECTIVE IN THE TREATMENT OF TUMORS AND OTHER  
TITLE OF INVENTION: CONDITIONS REQUIRING THE REMOVAL OR DESTRUCTION OF  
TITLE OF INVENTION: CELLS  
FILE REFERENCE: 59003.000008  
CURRENT APPLICATION NUMBER: US/10/198, 070  
CURRENT FILING DATE: 2002-07-19  
PRIOR APPLICATION NUMBER: 60/306,161  
PRIOR FILING DATE: 2001-07-19  
PRIOR APPLICATION NUMBER: 60/306,150  
PRIOR FILING DATE: 2001-07-19

PRIOR APPLICATION NUMBER: 60/331,477  
PRIOR FILING DATE: 2001-11-16  
NUMBER OF SEQ ID NOS: 125  
SOFTWARE: Patentln Ver. 2.1  
SEQ ID NO 67  
LENGTH: 392  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-198-070-67

Query Match 41.9%; Score 820; DB 9; Length 392;  
Best Local Similarity 43.1%; Pred. No. 1.3e-62;  
Matches 169; Conservative 80; Mismatches 131; Indels 12; Gaps 6;

OY 1 MASIAANAEFCNLFREMDNONGVNFSSLSLFAALALVRLGADDSISQIDKLLHV 60  
DB 1 MGSISTANVEFCIDVFEKLSNNIGDINFSSLSLALSLVLLGARGETAOLEKVLHF 60  
OY 61 -----NTASGYGNSN-SOSG-LOSOLKRVESDINASHKDYLSIVGLFAEKYGGPHKD 113  
DB 61 SHYVDSLKPFKDSQKSGRHSFGEVSQJNODSCTLSIANRLVGTKTMATHQ 120  
OY 114 YIECAERLYDAKERVDFTHLEDTRRNKKNVENETHGRIKINVGIGISSAVMYLVN 173  
DB 121 YLSCSEKMYQARLQTVDFEQSTETRTKINAMWENKNGKVANLFKSTIDPSSVAVLVN 180  
OY 174 AVYFKGKQSAFTSETINCHFKSPKSGKAVAMHDEKFNLSVIDPSSKILELY-N 232  
DB 181 TTYFGGROKMFQYREVKSPFQISEGNVYEMMYOIGTFKLAVEREPOQVLELYPVN 240  
OY 233 GGIMNYVLLPEN--DLSEIENKLFONLMEWTNPRRTSKYEVFFPOFKIEKNYEMKQY 290  
DB 241 NKLSMITLLPVGIANLQIEKQNSGFHEHTSSNMREVEVHLRFKILEITYELNSL 300  
OY 291 LRALGLDIFDESKADLSGASGRLYISRMHKSLEYTEEGTEATAGSNIVKQLP 350  
DB 301 LKPLGVTDLFNOVKADLSGMSPTKGLYLSKAIRKSYLDVSEGEAAATGDSIAVKSPL 360  
OY 351 OSTLFRAHDFPLFYIR--KDDILFSGKVSQP 380  
DB 361 MRAQFKANHPFLFFIRHTHTNTLLFCGLASP 392

RESULT 5  
US-10-198-070-70  
Sequence 70, Application US/10198070  
Publication No. US20030109437A1  
GENERAL INFORMATION:  
APPLICANT: AVERBACK, PAUL  
APPLICANT: GEMMELL, JACK  
TITLE OF INVENTION: PEPTIDES EFFECTIVE IN THE TREATMENT OF TUMORS AND OTHER  
TITLE OF INVENTION: CONDITIONS REQUIRING THE REMOVAL OR DESTRUCTION OF  
TITLE OF INVENTION: CELLS  
FILE REFERENCE: 59003.000008  
CURRENT APPLICATION NUMBER: US/10/198, 070  
CURRENT FILING DATE: 2002-07-19  
PRIOR APPLICATION NUMBER: 60/306,161  
PRIOR FILING DATE: 2001-07-19  
PRIOR APPLICATION NUMBER: 60/306,150  
PRIOR FILING DATE: 2001-07-19  
PRIOR APPLICATION NUMBER: 60/331,477  
PRIOR FILING DATE: 2001-11-16  
NUMBER OF SEQ ID NOS: 125  
SOFTWARE: Patentln Ver. 2.1  
SEQ ID NO 70  
LENGTH: 392  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-198-070-70

Query Match 41.8%; Score 819; DB 9; Length 392;  
Best Local Similarity 43.1%; Pred. No. 1.5e-62;  
Matches 169; Conservative 80; Mismatches 131; Indels 12; Gaps 6;

QY 1 MASLAANAEFCNLFREDDNOGNGVFFSSLSLFAALALVRLGAODDSLQIDKLHV 60  
 Db 1 MGSLSANVAFCDVFEKLNSNNIGNIFFSSLSLALYALSMVLLGARGETAEQLEVLHF 60  
 QY 61 -----NTASGYGNSSN-SQSG-LQSOLKRVFSDINASHKDYDLSYNGLEFAEKVYGFHKD 113  
 Db 61 SHTVDSLKPGFKSPQAGRIHSEGVFEFQINPDNSCTLSIANRLYGTMTAFHQ 120  
 QY 114 YIICAEKLYDAKERVDFNHLDTNRINKWVENETHGKIKNVIGEGISSAVAVLVN 173  
 Db 121 YLSCSEKWOARLQOTVDFDQSTETKRMIAVNTKNGVAVNLFGKSTIDPSSVAVLVN 180  
 QY 174 AYVEFKWQSAFETKSETINCHFPKSCGKAVAMHQRKPNLSVIEDPSMKILELRV-N 232  
 Db 181 IIVFKGQOROKFQVRETVKSPFQLSGKNVYEMMYQIGTFKLAFAVEKPMQVLELRYN 240  
 QY 233 GGIMMYVLLPEN--DLSEIENKLTFOINLMTNPRMTSKYVEVFPQFIKRYENKQY 230  
 Db 241 NKLSTMLILPVGIANLKOIEKQNSGTFHEWTSNNMEREVEVHLPRFKLEIKYEELNSL 300  
 QY 291 LRALGLKDIPESKADLSGIASGRILYISRMHKSYLEVTEEGTEATAATGSNIVEKOLP 350  
 Db 301 LKRLGTYDLFNQYKADLSGMSPTKGLYLSKAIHKSILYDSEEGTEAATAATGDSIAVKSLP 360  
 QY 351 QSTLFRADHPFLFVIR--KDDIILFSGKVSCLP 380  
 Db 361 MRAQFKNHPLFFIRHTHTNTILFCGKLASP 392

## RESULT 6

US-10-012-896-946  
 ; Sequence 946, Application US/10012896  
 ; Publication No. US20020183251A1

## GENERAL INFORMATION:

APPLICANT: Xu, Jiangchun  
 APPLICANT: Dillon, Davin C.  
 APPLICANT: Mitcham, Jennifer L.  
 APPLICANT: Harlocker, Susan L.  
 APPLICANT: Jiang, Yugu  
 APPLICANT: Kalos, Michael D.  
 APPLICANT: Retter, Marc W.  
 APPLICANT: Stolck, John A.  
 APPLICANT: Day, Craig H.  
 APPLICANT: Vedvick, Thomas S.  
 APPLICANT: Carter, Derrick  
 APPLICANT: Li, Samuel X.  
 APPLICANT: Wang, Aijun  
 APPLICANT: Skeiky, Yasir A.W.  
 APPLICANT: Hepler, William T.  
 APPLICANT: Henderson, Robert A.  
 APPLICANT: Hurai, John  
 APPLICANT: McNeill, Patricia D.  
 APPLICANT: Houghton, Raymond L.  
 APPLICANT: Vinals de Bassols, Carlota  
 APPLICANT: Foy, Teresa  
 APPLICANT: Fanger, Gary R.  
 APPLICANT: Mantanabe, Yoshihiro  
 APPLICANT: Meagher, Madeleine Joy  
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
 TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER  
 FILE REFERENCE: 210121.427C27  
 CURRENT APPLICATION NUMBER: US/10/012,896  
 CURRENT FILING DATE: 2001-12-10  
 NUMBER OF SEQ ID NOS: 1011  
 SOFTWARE: FastSeq for Windows Version 3.0  
 SEQ ID NO 946  
 LENGTH: 379  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-10-012-896-946

Query Match

40.7%; Score 796.5; DB 9; Length 379;

Best Local Similarity 42.6%; Pred. No. 1,36-60;  
 Matches 168; Conservative 77; Mismatches 120; Indels 29; Gaps 8;  
 QY 1 MASLAANAEFCNLFREDDNOGNGVFFSSLSLFAALALVRLGAODDSLQIDKLHV 60  
 Db 1 MGSLSANVAFCDVFEKLNSNNIGNIFFSSLSLALYALSMVLLGARGETAEQLEVLHF 60  
 QY 61 -----NTASGYGNSSN-SQSG-LQSOLKRVFSDINASHKDYDLSYNGLEFAEKVYGFHKD 113  
 Db 61 SHTVDSLKPGFKSPQAGRIHSEGVFEFQINPDNSCTLSIANRLYGTMTAFHQ 120  
 QY 114 YIICAEKLYDAKERVDFNHLDTNRINKWVENETHGKIKNVIGEGISSAVAVLVN 173  
 Db 121 YLSCSEKWOARLQOTVDFDQSTETKRMIAVNTKNGVAVNLFGKSTIDPSSVAVLVN 180  
 QY 174 AYVEFKWQSAFETKSETINCHFPKSCGK--AVAMHQRKPNLSVIDPSMKILELRV 231  
 Db 181 IIVFKGQOROKF-----QGNVTVEMMYQIGTFKLAFAVEKPMQVLELRY 225  
 QY 233 -NGIMMYVLLPEN--DLSEIENKLTFOINLMTNPRMTSKYVEVFPQFIKRYENK 288  
 Db 226 VNKLSMILILPVGIANLKOIEKQNSGTFHEWTSNNMEREVEVHLPRFKLEIKYEELN 285  
 QY 289 QYLRALGLKDIPESKADLSGIASGRILYISRMHKSYLEVTEEGTEATAATGSNIVEKQ 348  
 Db 286 SLKPLGVYDLFNQYKADLSGMSPTKGLYLSKAIHKSILYDSEEGTEAATAATGDSIAVKS 345  
 QY 349 LPQSTLFRADHPFLFVIR--KDDIILFSGKVSCLP 380  
 Db 346 LPMRAQFKNHPLFFIRHTHTNTILFCGKLASP 379

## RESULT 7

US-09-895-814-946  
 ; Sequence 946, Application US/09895814  
 ; Publication No. US20020193296A1

## GENERAL INFORMATION:

APPLICANT: Xu, Jiangchun  
 APPLICANT: Dillon, Davin C.  
 APPLICANT: Mitcham, Jennifer L.  
 APPLICANT: Harlocker, Susan L.  
 APPLICANT: Jiang, Yugu  
 APPLICANT: Kalos, Michael D.  
 APPLICANT: Retter, Marc W.  
 APPLICANT: Stolck, John A.  
 APPLICANT: Day, Craig H.  
 APPLICANT: Vedvick, Thomas S.  
 APPLICANT: Carter, Derrick  
 APPLICANT: Li, Samuel X.  
 APPLICANT: Wang, Aijun  
 APPLICANT: Skeiky, Yasir A.W.  
 APPLICANT: Hepler, William T.  
 APPLICANT: Henderson, Robert A.  
 APPLICANT: Hurai, John  
 APPLICANT: McNeill, Patricia D.  
 APPLICANT: Houghton, Raymond L.  
 APPLICANT: Vinals de Bassols, Carlota  
 APPLICANT: Foy, Teresa  
 APPLICANT: Fanger, Gary R.  
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
 TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER  
 FILE REFERENCE: 210121.427C26  
 CURRENT APPLICATION NUMBER: US/09/895,814  
 CURRENT FILING DATE: 2001-06-29  
 NUMBER OF SEQ ID NOS: 990  
 SOFTWARE: FastSeq for Windows Version 3.0  
 SEQ ID NO 946  
 LENGTH: 379  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-09-895-814-946

Query Match

40.7%; Score 796.5; DB 9; Length 379;

Best Local Similarity 42.68; Pred. No. 1.3e-60;  
Matches 168; Conservative 77; Mismatches 120; Indels 29; Gaps 8.

```

OY      1 MASLAAAANEEFNLEFREDDNOGNGNSNFSSLSLEPAALVYLGADODSLIDKLVH 60
Db      1 MGSLSTANVERCLDYFKELNSNNIGDNIFSSLSLTYALSMVLLGRKGETAEOLDKVLHF 60
OY      61 -----NTASGYGNSSN-SQSG-LQSLQKRVFSDINASHHDYDLSTYNGLSFAEKVYCGPHKD 113
Db      61 SHTVDSLARGFNDSPKCSQAGRIHSEFGEVFEISQINGPDSNCTLSIANRLTYGTMFAFHQ 120
OY      114 YIECEKLYDAKVERDYFTNLLEDTRRNINKVVENETBCKIKNVIGEGLSSAAVNLVN 173
Db      121 YLSCSEKRYOARLQTVDEQSTEEFRKMTNAVVENTNKVNLJFKSTIDPSSVYMLVN 180
OY      174 AVYFPGKQMSAFTKSETINCHFKSPKSCSG--AAVAMHDERKNLSVIEBPSMKILLEYR 231
Db      181 TYIFGGQONKF-----OGKNVTVEMVYIGTFKLVKVEPQOVLELPY 225
OY      232 -NGCINMYVLLPEN--DLSEIENKLTFOULMEVNTNRRMTSKYVEVFPQFIIEKVEMK 288
Db      226 VNNKSLMTILLPVGTANLKOIEKOLNSGFHEMTSSNNMREVEVHLRFKLEIKELYELN 285
OY      289 QYLRALGLKIDFDESKADLSGIASGRLYISRMHKSYLEVTEEGTEATPANSNIVEKO 348
Db      286 SILKPLGVTDLFNQVKADISGMSPTKGLTSLKAHKSXYLDVSEEGTEAANAATGDSIAVKS 345
OY      349 LPQSTLFRADHPFLFVIR--KDDILFSGKVCSP 380
Db      346 LPMRAQFPAHNPFLFIRHTHTNTNLTFCCKLSP 379

```

RESULT 8  
US-09-823-187-2

\* Sequence 2, Application US/09823187  
 \* Publication No. US2003009692A1  
 \* GENERAL INFORMATION:  
 \* APPLICANT: Burgess, Catherine  
 \* APPLICANT: Gusev, Vladimir Y  
 \* APPLICANT: Liu, Xiaohong  
 \* APPLICANT: Majumder, Kumud  
 \* APPLICANT: Patilgaru, Muralidhar  
 \* APPLICANT: Patursajan, Meera  
 \* APPLICANT: Shukrets, Richard A  
 \* APPLICANT: Spaderna, Steven K  
 \* APPLICANT: Spytek, Kimberly  
 \* APPLICANT: Tausler, Raymond J  
 \* TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME  
 \* FILE REFERENCE: 15966-745  
 \* CURRENT APPLICATION NUMBER: US/09/823,187  
 \* CURRENT FILING DATE: 2001-03-29  
 \* PRIOR APPLICATION NUMBER: 60/193,339  
 \* PRIOR FILING DATE: 2000-03-30  
 \* PRIOR APPLICATION NUMBER: 60/193,205  
 \* PRIOR FILING DATE: 2000-03-30  
 \* PRIOR APPLICATION NUMBER: 60/195,343  
 \* PRIOR FILING DATE: 2000-04-05  
 \* PRIOR APPLICATION NUMBER: 60/195,088  
 \* PRIOR FILING DATE: 2000-04-06  
 \* PRIOR APPLICATION NUMBER: 60/195,005  
 \* PRIOR FILING DATE: 2000-04-06  
 \* PRIOR APPLICATION NUMBER: 60/195,792  
 \* PRIOR FILING DATE: 2000-04-10  
 \* PRIOR APPLICATION NUMBER: 60/196,556  
 \* PRIOR FILING DATE: 2000-04-11  
 \* PRIOR APPLICATION NUMBER: 60/197,081  
 \* PRIOR FILING DATE: 2000-04-13  
 \* PRIOR APPLICATION NUMBER: 60/197,525  
 \* PRIOR FILING DATE: 2000-04-14  
 \* PRIOR APPLICATION NUMBER: 60/197,087  
 \* PRIOR FILING DATE: 2000-04-14  
 \* NUMBER OF SEQ ID NOS: 103  
 \* SOFTWARE: PatentIn Ver. 2.1

```

; SEQ ID NO 2
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-823-187-2

```

Query Match	39.4%;	Score 771.5;	DB 9;	Length 395;
Best Local Similarity	40.3%;	Pred. No. 2e-58;		
Matches 159;	Conservative 89;	Mismatches 132;	Indels 15;	Gaps 5

```

0Y 1 AASIIAAANAEEPCENFERREDDNOGNGNVEFSSLSIPAALALVYLGADODBSIQIDKLLH 60
Db 1 MDLSLTAMTKFCFEDLPÖEIGKDDRHNNTFFSPULSLSAAGMRVLRGRSDSAHOIDEVLRH 60
QY 61 NTASGY-----GNSSNSOSGLQ5-QUKRPFSDINMSHKDYDLSIVGLGFAEKRYVGHKDY 114
Db 61 NEFSGNESKEPAGSLNNESGLVSCYFGQLLSKDKTRKTDYTLISARLVGEQFPCQEX 120
QY 115 IECAEKLIDAKERYVDFTNHLIEDTRRNINKWYENBETHGKIKVYIGEGGJSSAVMYLVNA 174
Db 121 LDGVIOFPHYTTIESYDFQKNPEKSRQEIINFWECCOQGIKDFLSDADINAEVTLVYNA 180
QY 175 VYERKMKOSAFKSEFINCHFKSPKCSGKAAYVAMHMERPNLVSVIDPSMKIITELNY-NG 233
Db 181 VYEFKAMEYIEPHENTVDAPEFCLNOENENKSVKMMTKGLYRIGFIEVKAQIILEMRTYKG 240
QY 234 GINMYVLLPREND-----LSEIENKLTLPONIMEMTNPRIKMTSKRYVEVFFQFKIEKRYEM 287
Db 241 KLSMVLPLPSHSHKDNMLKGLEELERKTTYERKMYAMSSSENNSESVYLSPPRFTLLDSYDL 3000
QY 288 KOYLRALGLKIDIFDESKADLSGIASAGRLYISBRMHKSYIEVTEBGETEATATAGSNIVEK 347
Db 301 NSILODMGTTIDFDETRADLTGSI5PSPNYLSKTIHKTEVEVDENGT0AAAGAVASER 3600
QY 348 QLP0STLFRADHPFLFYVR--KDDIILFSGKVSXP 380
Db 361 SLRSWVEFNAHPLEFIRHNKTQYTLFLFGKVSXP 395

```

RESULT 9  
US-09-823-187-29

```

? Sequence 29, Application US//09823187
? Publication No. US20030096952A1
? GENERAL INFORMATION:
? APPLICANT: Burgess, Catherine
? APPLICANT: Gusev, Vladimir Y
? APPLICANT: Liu, Xiaohong
? APPLICANT: Majumder, Kumud
? APPLICANT: Padigar, Muralidhar
? APPLICANT: Patturajan, Meera
? APPLICANT: Shinkets, Richard A
? APPLICANT: Spaderna, Steven K
? APPLICANT: Szytek, Kimberly
? APPLICANT: Tappier, Raymond J
? TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
? FILE REFERENCE: 15966-745
? CURRENT APPLICATION NUMBER: US/09/823,187
? CURRENT FILING DATE: 2001-01-29
? PRIOR APPLICATION NUMBER: 60/193,339
? PRIOR FILING DATE: 2000-03-30
? PRIOR APPLICATION NUMBER: 60/193,205
? PRIOR FILING DATE: 2000-03-30
? PRIOR APPLICATION NUMBER: 60/195,343
? PRIOR FILING DATE: 2000-04-05
? PRIOR APPLICATION NUMBER: 60/195,088
? PRIOR FILING DATE: 2000-04-06
? PRIOR APPLICATION NUMBER: 60/195,005
? PRIOR FILING DATE: 2000-04-06
? PRIOR APPLICATION NUMBER: 60/195,792
? PRIOR FILING DATE: 2000-04-10
? PRIOR APPLICATION NUMBER: 60/196,556
? PRIOR FILING DATE: 2000-04-11
? PRIOR APPLICATION NUMBER: 60/197,081

```

```

RESULT 10
US-09-823-187-30
: Sequence 30. Application US/09823187
: Publication NO. US20030096952A1
: GENERAL INFORMATION:
: APPLICANT: Burgess, Catherine
: APPLICANT: Gusev, Vladimir Y
: APPLICANT: Liu, Xiaohong
: APPLICANT: Majumder, Rumud
: APPLICANT: Padigaru, Muralidhar
: APPLICANT: Patlurajan, Meera
: APPLICANT: Shimkets, Richard A
: APPLICANT: Spaderna, Steven K
: APPLICANT: Spytek, Kimberly
: APPLICANT: Taupier, Raymond J
: TITLE OR INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
: FILE REFERENCE: 15966-745
: CURRENT APPLICATION NUMBER: US/09/823,187
: CURRENT FILING DATE: 2001-03-29
: PRIOR APPLICATION NUMBER: 60/193,339
: PRIOR FILING DATE: 2000-03-30
: PRIOR APPLICATION NUMBER: 60/193,205
: PRIOR FILING DATE: 2000-03-30
: PRIOR APPLICATION NUMBER: 60/195,343
: PRIOR FILING DATE: 2000-04-05
: PRIOR APPLICATION NUMBER: 60/195,088
: PRIOR FILING DATE: 2000-04-06

```

RESULT 11  
US-09-823-187-27  
Sequence 27, Application US/09823187  
Publication No. US20030096952A1  
GENERAL INFORMATION:  
APPLICANT: Burgees, Catherine  
APPLICANT: Gusev, Vladimir Y  
APPLICANT: Liu, Xiaohong  
APPLICANT: Majumder, Kumud  
APPLICANT: Padigar, Muralidhar  
APPLICANT: Patnurajan, Meera  
APPLICANT: Shinkets, Richard A  
APPLICANT: Spaderna, Steven K  
APPLICANT: Szytek, Kimberly  
APPLICANT: Tausler, Raymond J  
TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME  
FILE REFERENCE: 15966-745  
CURRENT APPLICATION NUMBER: US/09/823,187  
CURRENT FILING DATE: 2001-03-29  
PRIOR APPLICATION NUMBER: 60/193,339

```

; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 60/193,205
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 60/195,343
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: 60/195,088
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/195,005
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/195,792
; PRIOR FILING DATE: 2000-04-10
; PRIOR APPLICATION NUMBER: 60/196,556
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: 60/197,081
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: 60/197,525
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/197,087
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 27
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-823-187-27

```

```

Query Match      38.2% Score 748; DB 9; Length 390;
Best Local Similarity 40.6%; Pred. No. 2e-56;
Matches 160; Conservative 84; Mismatches 134; Indels 14; Gaps 8;

```

```

QY 1 MASIAAANAEFCNLFREMDNGNGVNFSSLSLFAALALVRLGADDSLSQTDKLHY 60
D 1 MNSLSEANTFEMFDLFOQFRKSKEN-NIFSPISITSLALGMVLGADNDAQOISKVLHF 59
QY 61 -----NTASGYGNSSNSQSG-LOSOLKRVFSDINASHKDYDLSTVNGLFAEKYVGFHKDY 114
D 60 DQYVENTEKAATYHYDRSGNVHNGQKLLTEFNKSDAELKANKLFGEKTYQFLOEY 119
QY 115 IECAEKLYDAKVERVDFTNLEDTRRINKNVENETHGKIKNYIGEGISSAAVVLVNA 174
D 120 LDAIKKRYQSVESSTDFANAPAESRRKINSWVESQTEKIKNLFPPDGTIGNDTTLVLYNA 179
QY 175 VYRKGKQSAFTSETINCHFKSPKCGKAVAMMHORKNLSVIEDPSKILELRNG- 233
D 180 IYRKGQENKFKKENTKEEFWPKRNTYKSVOMRQNSFNFALEEDVQAKVLEIPYK 239
QY 234 GINMYVLLPE--NDLSEIENKLTFOIMLEMTNPRMTSKYVEVEFFPOFKIEKNYEMQYL 291
D 240 DLSMIVLLPNEIDGLQLEKLEKLTAEKLEMTSLQNMRETCVDLHLPRFKMEESYDKDTL 299
QY 292 RALGLKDIFDESKADLSGIASGRLYISRMHKSYLEVTEEGTEATATGSNIYVKOLPQ 351
D 300 RTMGWNVIFN-GDADLSGTMWSHGLSVKYLHKAFAVEVEEGVEAAAATAVAVVEELSSPS 358
QY 352 ST-LFRADHPFLFYIR--KDDITLFSGKVCSP 380
D 359 TNEEFCNHPFLFPIRQNKTNLSILEYGRFSSP 390

```

```

RESULT 12
US-09-823-187-28
; Sequence 28, Application US/09823187
; Publication No. US20030096952A1
; GENERAL INFORMATION:
; APPLICANT: Burgess, Catherine
; APPLICANT: Gusev, Vladimir Y
; APPLICANT: Liu, Xiaohong
; APPLICANT: Majumder, Kumud
; APPLICANT: Padigaru, Muralidhar
; APPLICANT: Patlurajan, Meera
; APPLICANT: Shinkets, Richard A
; APPLICANT: Spaderna, Steven K

```

```

; APPLICANT: Spytek, Kimberly
; APPLICANT: Taulier, Raymond J
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 15966-745
; CURRENT FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: US/09/823,187
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 60/193,339
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 60/193,205
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 60/195,343
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: 60/195,088
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/195,005
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/195,792
; PRIOR FILING DATE: 2000-04-10
; PRIOR APPLICATION NUMBER: 60/196,556
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: 60/197,081
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: 60/197,525
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/197,087
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-823-187-28

```

```

Query Match      37.9% Score 743; DB 9; Length 390;
Best Local Similarity 40.6%; Pred. No. 5.5e-56;
Matches 159; Conservative 84; Mismatches 135; Indels 14; Gaps 8;

```

```

QY 1 MASIAAANAEFCNLFREMDNGNGVNFSSLSLFAALALVRLGADDSLSQTDKLHY 60
D 1 MNSLSEANTFEMFDLFOQFRKSKEN-NIFSPISITSLALGMVLGADNDAQOISKVLHF 59
QY 61 -----NTASGYGNSSNSQSG-LOSOLKRVFSDINASHKDYDLSTVNGLFAEKYVGFHKDY 114
D 60 DQYVENTEKAATYHYDRSGNVHNGQKLLTEFNKSDAELKANKLFGEKTYQFLOEY 119
QY 115 IECAEKLYDAKVERVDFTNLEDTRRINKNVENETHGKIKNYIGEGISSAAVVLVNA 174
D 120 LDAIKKRYQSVESSTDFANAPAESRRKINSWVESQTEKIKNLFPPDGTIGNDTTLVLYNA 179
QY 175 VYRKGKQSAFTSETINCHFKSPKCGKAVAMMHORKNLSVIEDPSKILELRNG- 233
D 180 IYRKGQENKFKKENTKEEFWPKRNTYKSVOMRQNSFNFALEEDVQAKVLEIPYK 239
QY 234 GINMYVLLPE--NDLSEIENKLTFOIMLEMTNPRMTSKYVEVEFFPOFKIEKNYEMQYL 291
D 240 DLSMIVLLPNEIDGLQLEKLEKLTAEKLEMTSLQNMRETCVDLHLPRFKMEESYDKDTL 299
QY 292 RALGLKDIFDESKADLSGIASGRLYISRMHKSYLEVTEEGTEATATGSNIYVKOLPQ 351
D 300 RTMGWNVIFN-GDADLSGTMWSHGLSVKYLHKAFAVEVEEGVEAAAATAVAVVEELSSPS 358
QY 352 ST-LFRADHPFLFYIR--KDDITLFSGKVCSP 380
D 359 TNEEFCNHPFLFPIRQNKTNLSILEYGRFSSP 390

```

```

RESULT 13
US-08-731-566-2
; Sequence 2, Application US/08731566
; Publication No. US20030072752A1
; GENERAL INFORMATION:
; APPLICANT: Toshiniko Hibino et al.

```

TITLE OF INVENTION: TWO NOVEL GENES FROM PSORIASIS EPIDERMIS:  
TITLE OF INVENTION: PSORIASIS TYPE I AND PSORIASIS TYPE II  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD, LLP  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1775  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/731,566  
FILING DATE: 16-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/013,755  
FILING DATE: 20-MAR-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/005,679  
FILING DATE: 17-OCT-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Louis Meyers  
REGISTRATION NUMBER: 35,965  
REFERENCE/DOCKET NUMBER: MGP-040CP  
TELEPHONE: (617)227-7400  
TELEFAX: (617)227-5941  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 390 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-731-566-2

Query Match 37.8%; Score 740; DB 7; Length 390;

Best Local Similarity 41.0%; Pred. No. 1e-55; Indels 26; Gaps 10;

Matches 103; Conservative 81; Mismatches 128; Indels 26; Gaps 10;

QY 1 MASLAANAEPFCNLEFREMDDNGNGNVEFSSLSLEPAALALVRLGAQDSSLQIDKLHY 60  
1 MMSLSANTKPMEDLFQFRKSKEN-NITYSPISIALGMVLLGAKGNTAQDICKVLFH 59  
QY 61 -----NTASGYGNSSNSOG-LOSOLKRVPSDINASHKDYDLSYVGLFAEKVYGFHKDY 114  
60 DQVTEHTTKAATYHYDRSGDVHNOFQKLTTEFNKSTDAVELKANKLGEKTYLFLQCY 119  
QY 115 IEOAEKLYDAKVERVFTNHLDETRNINKWVENETHGKIKNYIGGGTSSSAVMVLVA 174  
120 LNAIKFVOTSVESVDFAAPESRKKINSWESOTNEKIKNLPISGINSNTLVVLA 179  
QY 175 VYFKGMOAFTKSETINCHFKSPKSGRAVAMHOERFNLVIEDPSKILRLRYNG- 233  
180 IYFGKMEKFKNEDEKFKPMNKTYKSIGMMQRYTSFHFASLEDYQAVALEITYKGR 239  
QY 234 GINMYVLPE--NDLSEINLKLTFQNLMEFTNRRMTSKYVEYFPOFKIEKNYKQYL 291  
240 DLSMIVLLPNEIDGLQRLLEEELTAELKLMETSLQNMRRETRVDLHLRFKVEESYDLKDTL 299  
QY 292 RALGKIDFDESKADISGASGRLLYSRMHKSYLEVTEEGEATNAAT-----GSNIYE 346  
300 RTMGAVDITN-GDADISGMTSGRLVSLVLAHFAVEYEEGAALAAVVGFGSS--- 355  
QY 347 KOLPOST--LFRADHPFLFVIR--KDDIILFSGVSCP 380  
356 ---PSTNEFHCHNHPFLFFIRONKNTSILFYGRSSP 390  
Db

RESULT 14

US-10-012-896-947  
Sequence 947, Application US/10012896  
Publication No. US20020183251A1  
GENERAL INFORMATION:  
APPLICANT: Xu, Jiangchun  
APPLICANT: Dillon, Davin C.  
APPLICANT: Mitcham, Jennifer L.  
APPLICANT: Harlocker, Susan L.  
APPLICANT: Jiang, Yugu  
APPLICANT: Kalos, Michael D.  
APPLICANT: Retter, Marc W.  
APPLICANT: Stolk, John A.  
APPLICANT: Day, Craig H.  
APPLICANT: Vedvick, Thomas S.  
APPLICANT: Carter, Darrick  
APPLICANT: Li, Samuel X.  
APPLICANT: Wang, Aijun  
APPLICANT: Skelky, Yasir A.W.  
APPLICANT: Hepler, William T.  
APPLICANT: Henderson, Robert A.  
APPLICANT: Hural, John  
APPLICANT: McNeill, Patricia D.  
APPLICANT: Houghton, Raymond L.  
APPLICANT: Vinals de Bassols, Carlota  
APPLICANT: Foy, Teresa  
APPLICANT: Fanger, Gary R.  
APPLICANT: Mantabe, Yoshihiro  
APPLICANT: Meagher, Madeleine Joy  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
FILE REFERENCE: 210121.427C27  
CURRENT APPLICATION NUMBER: US/10/012,896  
CURRENT FILING DATE: 2001-12-10  
NUMBER OF SEQ ID NOS: 1011  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 947  
LENGTH: 617  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: 90, 215, 225, 232, 233, 289, 309, 318, 322, 331, 332, 367,  
LOCATION: 369, 374, 378, 425, 430  
OTHER INFORMATION: Xaa = Any Amino Acid  
US-10-012-896-947

Query Match 37.7%; Score 738.5; DB 9; Length 617;

Best Local Similarity 28.8%; Pred. No. 2.6e-55; Indels 237; Gaps 11;

Matches 178; Conservative 81; Mismatches 121;

QY 1 MASLAANAEPFCNLEFREMDDNGNGNVEFSSLSLEPAALALVRLGAQDSSLQIDKLHY 60  
1 MMSLSANTKPMEDLFQFRKSKEN-NITYSPISIALGMVLLGAKGNTAQDICKVLFH 60  
QY 61 -----NTASGYGNSSNSOG-LOSOLKRVPSDINASHKDYDLSYVGLFAEKVYGFHKDY 113  
61 SHYVDSLKRGKFDSPKCSQAGRIHSEFGVXFQIOPDNCSTLSINRLYGRKTAFFHQ 120  
QY 114 YIECAEKLDAKVERVFTNHLDETRNINKWVENETHGKIKNYIGGGTSSSAVMVLVA 173  
121 YLSCSEKQWARIQYVDFEOSTEETRTKINAWENTKGNVNLFGKSTIDPSSVMVLVA 180  
QY 174 AYFKGMOAFTKSETIN- 192  
181 ATEFGKQWONKQOVRETVASPPOLSEVSLFSDSXOMLEDITIIKQFPRKMXFSENIGL 240  
QY 193 ----- 192  
241 GFCWFLLYFLQIFLIPLLSDNNFYHRAENWRLGILRFGSGRGNPFESXSLGLFEPYIL 300  
QY 193 -----CHP-----KSPK----- 199  
301 WLCSFAHAGYLCYFFFXHVSXGKIKKMIXXYIILPTKIMLAKNPDEVGFRPSYLYL 360  
Db

```

Qy 200 -----CS----- 201
Db 361 LEQSLAPKXLLNKKNGXPLOREVIYNLCSEFTFHAERVMOISVLKRVISTHICALTY 420
Qy 202 -----GK--AVAMHQRKFNLSVIEDPSMKILELRY-NGGINMVLLEN--D 245
Db 421 VSLXSPSSXOGKKNVTVMYQIGTFKLAFAKPEOMQVLELPYNNKLSMILLPVGIAN 480
Qy 246 LSEIENKLTFOINLMEINPRMRTSKYVEVFPQRIEKNYEMKOYLALGLKIDFDESKA 305
Db 481 LKQIEKQINSCTFHEWTSSSNMEREVEVHLPRFKLETKYELNSLKSLSGLVTDLPNOYKA 540
Qy 306 DLGSIASGRLYISRMHKSYTEVEBTEATAGSINIVEKOLPOSTLFRADHPFLVI 365
Db 541 DLGSMSPYKGLYLKAIHKSYLVDSECTEAAAATGDSIAVKSLLPMAQFANHPFLFI 600
Qy 366 R-KDITLFSGKVSXP 380
Db 601 RHTHTNTILFCGKLASP 617

RESULT 15
US-09-895-814-947
: Sequence 947, Application US/09895814
: Publication No. US20020193296A1
: GENERAL INFORMATION:
: APPLICANT: Xu, Jiangchun
: APPLICANT: Dillon, Davin C.
: APPLICANT: Mitchem, Jennifer L.
: APPLICANT: Harlocker, Susan L.
: APPLICANT: Jiang, Yugu
: APPLICANT: Kalos, Michael D.
: APPLICANT: Retter, Marc W.
: APPLICANT: Stoik, John A.
: APPLICANT: Day, Craig H.
: APPLICANT: Vedvick, Thomas S.
: APPLICANT: Carter, Darick
: APPLICANT: Li, Samuel X.
: APPLICANT: Wang, Aljun
: APPLICANT: Skeiky, Yasir A.W.
: APPLICANT: Hepler, William T.
: APPLICANT: Henderson, Robert A.
: APPLICANT: Hurral, John
: APPLICANT: McNeill, Patricia D.
: APPLICANT: Houghton, Raymond L.
: APPLICANT: Vinals de Bassols, Carlota
: APPLICANT: Foy, Teresa
: APPLICANT: Fanger, Gary R.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
: FILE REFERENCE: 210121.427C26
: CURRENT APPLICATION NUMBER: US/09/895, 814
: NUMBER OF SEQ ID NOS: 990
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 947
: LENGTH: 617
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: variant
: LOCATION: (1)...(617)
: OTHER INFORMATION: Xaa - Any amino acid
US-09-895-814-947

Query Match 37.7%; Score 738.5; DB 9; Length 617;
Best Local Similarity 28.8%; Pred. NO. 2.6e-55;
Matches 178; Conservative 81; Mismatches 121; Indels 237; Gaps 11;
Qy 1 MASLAANAEPFCNLFRMDNCGNNGVFFSSLSLFAALALVRLGADDSLSQIDKLLHV 60
Db 1 MGSLSANVEFCLDVFELNNSNIGDNIFSSLSLSLVALSVLLGARGETAQLEKVLHF 60

```

```

Qy 61 -----NTASGYGNSSN-SOSG-LOSQLRVESDINASHKDYDLSTVNGLPFAEKYGFPHD 113
Db 61 SHYDLSLKPGRDSCQAGRHSEFVYXSOINOPSSCTLSIANLITGKTMAFMOQ 120
Qy 114 YIECAEKLYDAKVERVDFTNHLDRNINKNVENETGKIKNYIGEGISSAVMYLVN 173
Db 121 YLSCSEKMYQARLQVDFEQSTERTKTINMVEKKTNGKVANLFGKSTIDPSSVWLVN 180
Qy 174 AVERGKMQASAPTYSETIN----- 192
Db 181 AIYFQGMQNRFOYARETVKSPFOLSEVSLFSDSXOMLEDYIIIXGPRKMXFSENIQL 240
Qy 193 ----- 192
Db 241 GFCWFELLYLQIFLIFPLSDNNFYHRAPNRBLGILRSGCENPFESXRSLGLFFPYIL 300
Qy 193 -----CHP-----KSPK----- 199
Db 301 WLCSPAHHXGYLCYFEFFXRVSGKIKKMXIXXYILFLPTKIMLAKNPDVFGRPXYIL 360
Qy 200 -----CS----- 201
Db 361 LEQSLAPKXLLNKKNGXPLOREVIYNLCSEFTFHAERVMOISVLKRVISTHICALTY 420
Qy 202 -----GK--AVAMHQRKFNLSVIEDPSMKILELRY-NGGINMVLLEN--D 245
Db 421 VSLXSPSSXOGKKNVTVMYQIGTFKLAFAKPEOMQVLELPYNNKLSMILLPVGIAN 480
Qy 246 LSEIENKLTFOINLMEINPRMRTSKYVEVFPQRIEKNYEMKOYLALGLKIDFDESKA 305
Db 481 LKQIEKQINSCTFHEWTSSSNMEREVEVHLPRFKLETKYELNSLKSLSGLVTDLPNOYKA 540
Qy 306 DLGSIASGRLYISRMHKSYTEVEBTEATAGSINIVEKOLPOSTLFRADHPFLVI 365
Db 541 DLGSMSPYKGLYLKAIHKSYLVDSECTEAAAATGDSIAVKSLLPMAQFANHPFLFI 600
Qy 366 R-KDITLFSGKVSXP 380
Db 601 RHTHTNTILFCGKLASP 617

```

Search completed: July 11, 2003, 12:37:36  
Job time : 112.393 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:45:06 ; Search time 76.632 Seconds  
(without alignments)  
476.708 Million cell updates/sec

Title: US-10-091-442-34

Perfect score: 1959

Sequence: 1 MASLAANAEFCNLFREMD.....FLFVIRKDDILFSGKVSCE 380

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	752	38.4	390	2	I38201
2	743	37.9	390	2	I38202
3	739.5	37.7	379	2	A42421
4	722.5	36.9	379	2	S27383
5	715.5	36.5	391	2	JC7118
6	711	36.3	388	1	DYCH
7	699	35.7	378	2	S38962
8	664.5	33.9	415	2	A32853
9	662	33.8	374	2	A59273
10	659	33.6	386	1	OACH
11	658.5	33.6	383	2	S11433
12	658.5	33.6	397	2	I39184
13	649	33.1	378	2	A57488
14	648	33.1	376	1	A48681
15	630.5	32.2	415	2	S20047
16	623	31.8	376	2	B59273
17	622	31.8	416	2	S19896
18	592	30.2	465	2	I59611
19	590.5	30.1	375	2	A36888
20	578	29.5	431	1	JX0364
21	567	28.9	433	1	A61435
22	566	28.9	464	1	XHHU3
23	557	28.4	465	1	S28219
24	478.5	24.4	417	2	S19724
25	476.5	24.3	410	2	S70647
26	476	24.3	418	2	S23675
27	468.5	23.9	359	2	D88940
28	467	23.8	410	2	A45457
29	464.5	23.7	418	2	A53120

30	463.5	23.7	433	1	ITHUC	alpha-1-antichymot
31	462.5	23.6	418	2	JX0129	contrapsin precurs
32	461.5	23.6	397	2	T06183	serpin - barley
33	460	23.5	418	1	S31507	serine proteinase
34	450	23.0	408	2	S11320	serine proteinase
35	446.5	22.8	405	2	A39088	alpha-1-antiprotei
36	446.5	22.8	406	2	A39339	protein C inhibitor
37	444	22.7	366	2	T16119	hypothetical prote
38	438.5	22.4	413	2	JX0267	alpha-1-antiprotei
39	437	22.3	418	2	JH0494	alpha-1-antichymot
40	435.5	22.2	403	2	S08102	serine proteinase
41	434	22.2	416	2	B29131	kallikrein-binding
42	433.5	22.1	413	2	JX0154	alpha-1-antiprotei
43	433	22.1	232	1	DYCH	ovalbumin-related
44	432	22.1	479	1	S41066	heparin cofactor I
45	430	21.9	366	2	T25504	hypothetical prote

## ALIGNMENTS

RESULT 1  
I38201  
squamous cell carcinoma antigen 1 - human  
C:Species: Homo sapiens (man)  
C:Date: 23-Feb-1996 #sequence, revision 23-Feb-1996 #text, change 26-May-2000  
C:Accession: I38201, I38200, G01631, J70966, J70967  
R:Schneider, S.S.; Schick, C.; Fish, K.E.; Miller, E.; Pena, J.C.; Treter, S.D.; Hul, proc. Natl. Acad. Sci. U.S.A. 92, 3147-3151, 1995  
A:Title: A serine proteinase inhibitor locus at 18q21.3 contains a tandem duplication  
A:Reference number: I38200; MUID:95241462; PMID:7724531  
A:Accession: I38201  
A>Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-390 <SCCH1>  
A:Cross-references: EMBL:U19568; GB:U19558; NID:g1172085  
A:Accession: I38200  
A>Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-117 <SCH2>  
A:Cross-references: EMBL:U19562; NID:9897835; PIDN:AAA86316.1; PID:9897844  
R:Silverman, G.A.  
A:Reference number: G07968  
A:Accession: G01631  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-350, 'G', 352-390 <SHL>  
A:Cross-references: EMBL:U19556; NID:g1276435; PID:g1052869  
R:Sumitani, Y.; Kishi, F.; Sekiguchi, K.; Kato, H. Biochem. Biophys. Res. Commun. 181, 51-58, 1991  
A:Title: Squamous cell carcinoma antigen is a new member of the serine protease inhib  
A:Reference number: J70966; MUID:92068241; PMID:1958219  
A:Accession: J70966  
A:Molecule type: mRNA  
A:Residues: 1-350, 'G', 352-356, 'A', 358-390 <SUM1>  
A:Cross-references: GB:S66896; NID:g239551; PIDN:ABB20405.1; PID:g239552  
A:Accession: J70967  
A:Molecule type: protein  
A:Residues: 11-21, 231-237, 240-256, 303-325 <SUM2>  
C:Comment: This antigen probably acts as a proteinase inhibitor to modulate the host  
C:Genetics:  
A:Gene: GDB:SCCA1; SCC  
A:Cross-references: GDB:625364; OMIM:600517  
A:Map position: 18q21.3-18q21.3  
A:Introns: 55/3; 74/3; 117/3; 157/1; 204/3; 256/3  
C:Superfamily: antithrombin III  
C:Keywords: cysteine proteinase inhibitor; glycoprotein  
F:65/93,171,376/Binding site: carboxylate (Asn) (covalent) #status predicted  
F:354/Inhibitory site: Ser (cathepsin L) #status predicted

Query Match 38.4%; Score 752; DB 2; Length 390;  
Best Local Similarity 41.5%; Pred. No. 1.3e-45;

Matches 155; Conservative 81; Mismatches 126; Indels 26; Gaps 10;

QY 1 MASIAAANFEFCNLFREMDNOGNGNVEFSSLSLFAALALVRLGAQDSSLQIDKLLHV 60  
 1 MNSLSEANTKFMEDLQOEFKRSKEN-NIFESPISITSALGMVLLGAKDNTAQQISKVLHF 59  
 Db 61 -----NTASGYGSSNSOG-LOSQKRVFSDINASHKDYDSTVNGLFAEKVYGFHKDY 114  
 60 DQYENTTGTGAATYHYHDSRGVNHQFOKLTFENKSTDAVELKIANKLGEKTYLFLQEX 119  
 Db 115 IECAEKLVDKAYERVDFTNHLEDTNRNINKWENETHGRIKKNVIGEGISSSAVMVLVA 174  
 120 LDAIKRFYQTSVESTDFANAPESRRKINSWESQNEKIKNLPDGTIGNDTTLVANA 179  
 Db 175 VYFKGWSQAFKSEITNCHFKSPKCSGAVAMHBERKFNLSVIDPSMKILELRYNG- 233  
 180 IYKGGWKKFNEDRKEEFKFNKNTYKSIQMMRQYTSFHFASLSDVQAKVLEIYKCK 239  
 Db 234 GIMMYVLPE--NDLSEIENKLTFOULMEWTPRRRTSKYVEVFPOFKIEKVEKQYL 291  
 240 DLSMIYLLNEIDGLQLEKLEKLTAEKLEMTSLQNNRETRVDLHLPFRFVEESYDLKDL 299  
 QY 292 RALGLDIFDESKADLSGIASGRGLYISRMHKSYLEVTEEGTEATATAT--GSNIYE 346  
 300 RIMGWYDIFN-GDADLSGMTSHGLSVKYLKRAFEVTEEGEAAATAVVAFFSS--- 355  
 Db 347 KQLPOST--LFRADHPFLFVIR--KDDIILFSGKVS 380  
 356 ---PTSTNEEFCHNHPFLFIRQNKNTNSILFYGRFSSP 390

## RESULT 2

leupin precursor - human  
 N:Alternate names: protease inhibitor 11 (P11); squamous cell carcinoma antigen 2  
 C:Species: Homo sapiens (man)  
 C:Date: 23-Feb-1996 #sequence, revision 23-Feb-1996 #text, change 26-May-2000  
 C:Accession: I38202; S66675; S57522  
 R:Schneider, S.S.; Schick, C.; Fish, K.E.; Miller, E.; Pena, J.C.; Treter, S.D.; Hui, S.  
 Proc. Natl. Acad. Sci. U.S.A. 92, 3147-3151, 1995  
 A:Title: A serine proteinase inhibitor locus at 18q21.3 contains a tandem duplication of  
 A:Reference number: I38200; M0ID:95241462; PMID:7724531  
 A:Accession: I38202  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: DNA  
 A:Residues: 1-390 <SCN>  
 A:Cross-references: EMBL:U19576; GB:U19569; NID:g852466  
 R:Barnes, R.C.; Morrall, D.M.  
 FEBS Lett. 373, 61-65, 1995  
 A:Title: Identification of a novel human serpin gene: cloning sequencing and expression  
 A:Reference number: S66675; M0ID:96013887; PMID:7589435  
 A:Accession: S66675  
 A:Molecule type: mRNA  
 A:Residues: 7-351, 'V', 353-384 <BAR>  
 A:Cross-references: EMBL:X89015; NID:g887464; PIDN:CAA61420.1; PID:g887465  
 C:Genetics:  
 A:Gene: GDB:SCA2; P11  
 A:Cross-references: GDB:636556; OMIM:600518  
 A:Map position: 18q21.3-18q21.3  
 A:Introns: 55/3; 74/3; 117/3; 157/1; 204/3; 256/3  
 C:Superfamily: antithrombin III  
 C:Keywords: glycoprotein; serine proteinase inhibitor  
 F:1-24/Domain: signal sequence #status predicted <Stid>  
 F:23-384/Product: leupin #status predicted <MAT>  
 F:65,93/170,376/Binding site: carbonylrate (Asn) (covalent) #status predicted  
 F:354/Inhibitory site: Leu (unidentified proteinase) #status predicted

Query Match 37.9%; Score 743; DB 2; Length 390;  
 Best Local Similarity 40.6%; Pred. No. 5, 4e-45;  
 Matches 159; Conservative 84; Mismatches 135; Indels 14; Gaps 8;

QY 1 MASIAAANFEFCNLFREMDNOGNGNVEFSSLSLFAALALVRLGAQDSSLQIDKLLHV 60  
 1 MNSLSEANTKFMEDLQOEFKRSKEN-NIFESPISITSALGMVLLGAKDNTAQQISKVLHF 59

Db 1 MNSLSEANTKFMEDLQOEFKRSKEN-NIFESPISITSALGMVLLGAKDNTAQQISKVLHF 59  
 QY 61 -----NTASGYGSSNSOG-LOSQKRVFSDINASHKDYDSTVNGLFAEKVYGFHKDY 114  
 60 DQYENTTGTGAATYHYHDSRGVNHQFOKLTFENKSTDAVELKIANKLGEKTYLFLQEX 119  
 Db 115 IECAEKLVDKAYERVDFTNHLEDTNRNINKWENETHGRIKKNVIGEGISSSAVMVLVA 174  
 120 LDAIKRFYQTSVESTDFANAPESRRKINSWESQNEKIKNLPDGTIGNDTTLVANA 179  
 Db 175 VYFKGWSQAFKSEITNCHFKSPKCSGAVAMHBERKFNLSVIDPSMKILELRYNG- 233  
 180 IYKGGWKKFNEDRKEEFKFNKNTYKSIQMMRQYTSFHFASLSDVQAKVLEIYKCK 239  
 Db 234 GIMMYVLPE--NDLSEIENKLTFOULMEWTPRRRTSKYVEVFPOFKIEKVEKQYL 291  
 240 DLSMIYLLNEIDGLQLEKLEKLTAEKLEMTSLQNNRETRVDLHLPFRFVEESYDLKDL 299  
 QY 292 RALGLDIFDESKADLSGIASGRGLYISRMHKSYLEVTEEGTEATATAT--GSNIYE 346  
 300 RIMGWYDIFN-GDADLSGMTSHGLSVKYLKRAFEVTEEGEAAATAVVAFFELSSPS 358  
 Db 359 TNEEFCCNHPFLFIRQNKNTNSILFYGRFSSP 390

## RESULT 3

leukocyte elastase inhibitor - horse  
 N:Alternate names: plasminogen activator inhibitor-2 homolog  
 C:Species: Equus caballus (domestic horse)  
 C:Date: 04-Mar-1993 #sequence, revision 18-Nov-1994 #text, change 02-Jun-2000  
 C:Accession: A42421; A37276; S34062  
 R:Dubin, A.; Travis, J.; Enghild, J.J.; Potempa, J.  
 J. Biol. Chem. 267, 6576-6583, 1992  
 A:Title: Equine leukocyte elastase inhibitor. Primary structure and identification as  
 A:Reference number: A42421; M0ID:92202200; PMID:1551869  
 A:Accession: A42421  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 1-379 <DUB>  
 A:Cross-references: PIDN:AA021885.1; PID:g247842  
 A:Experimental source: leukocyte  
 A:Note: Sequence extracted from NCBI backbone (NCBIP:89849)  
 R:Dubin, A.; Travis, J.; Enghild, J.J.; Potempa, J.  
 submitted to the Protein Sequence Database, December 1991  
 A:Reference number: A37276  
 A:Accession: A37276  
 A:Molecule type: protein  
 A:Residues: 1-41, 'E', 43-325, 'VD', 326-379 <DUB>  
 R:Kordula, T.; Dubin, A.; Schoollink, H.; Koj, A.; Heinrich, P.C.; Rose-John, S.  
 Biochem. J. 293, 187-193, 1993  
 A:Title: Molecular cloning and expression of an intracellular serpin: an elastase inh  
 A:Reference number: S34062; M0ID:93319507; PMID:7687128  
 A:Accession: S34062  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-379 <KOR>  
 A:Cross-references: GB:M91161; NID:9164240; PIDN:AAA97513.1; PID:9164241  
 C:Superfamily: antithrombin III  
 C:Keywords: serine proteinase inhibitor

Query Match 37.7%; Score 739.5; DB 2; Length 379;  
 Best Local Similarity 38.0%; Pred. No. 9, 2e-45;  
 Matches 148; Conservative 85; Mismatches 137; Indels 19; Gaps 4;

QY 1 MASIAAANFEFCNLFREMDNOGNGNVEFSSLSLFAALALVRLGAQDSSLQIDKLLHV 60  
 1 MNSLSEANTKFMEDLQOEFKRSKEN-NIFESPISITSALGMVLLGAKDNTAQQISKVLHF 59  
 Db 61 -----NTASGYGSSNSOG-LOSQKRVFSDINASHKDYDSTVNGLFAEKVYGFHKDY 114  
 60 DQYENTTGTGAATYHYHDSRGVNHQFOKLTFENKSTDAVELKIANKLGEKTYLFLQEX 119  
 Db 115 IECAEKLVDKAYERVDFTNHLEDTNRNINKWENETHGRIKKNVIGEGISSSAVMVLVA 174  
 120 LDAIKRFYQTSVESTDFANAPESRRKINSWESQNEKIKNLPDGTIGNDTTLVANA 179  
 Db 175 VYFKGWSQAFKSEITNCHFKSPKCSGAVAMHBERKFNLSVIDPSMKILELRYNG- 233  
 180 IYKGGWKKFNEDRKEEFKFNKNTYKSIQMMRQYTSFHFASLSDVQAKVLEIYKCK 239  
 Db 234 GIMMYVLPE--NDLSEIENKLTFOULMEWTPRRRTSKYVEVFPOFKIEKVEKQYL 291  
 240 DLSMIYLLNEIDGLQLEKLEKLTAEKLEMTSLQNNRETRVDLHLPFRFVEESYDLKDL 299  
 QY 292 RALGLDIFDESKADLSGIASGRGLYISRMHKSYLEVTEEGTEATATAT--GSNIYE 346  
 300 RIMGWYDIFN-GDADLSGMTSHGLSVKYLKRAFEVTEEGEAAATAVVAFFELSSPS 358  
 Db 359 TNEEFCCNHPFLFIRQNKNTNSILFYGRFSSP 390

Db 61 DTN-----EDHSRFQSLNADINKGAPYLKLANRKYGEKTYNEFLADFLASTOK 110  
 QY 121 LYDAKVERVDFTNHLDETRRNINKWVENETHGKIKNYIGGGISSAVLVNAYVYKRG 180  
 Db 111 MYGAEIASYDFOQAPPEAKKEINEMVKGOTEGKPELVKGMVDMTKVLVNAIYKGN 170  
 QY 181 MOSAFKSETINCHFKSPKSGKAVAMHOERKFNLSVIDPSMKILELYNG-GIMMYV 239  
 Db 171 MOEFKMEATRDAPFRLNKOTKYKMMYQKKKFPYVYIEDLKCRVLELPYQCKELSMII 230  
 QY 240 LLEPN-----DSEIENKLTFLONLMEWTPRRMTSKYVEYFFPOFKIEKNYKQYLRA 293  
 Db 231 LRPDDIEDESTGLEKIEKQLTLEKLRMTKPEMLYLAEVVHLPRFKLEESYDLETSLAR 290  
 QY 294 LGLKDIDESKADLSGIASGRLYISRMHKSYTEVEEGEATAGTSNIVEKQLPOST 353  
 Db 291 LGVODLEFNKRGKADLSGSGARDLFVSKTIHSFVDLEEGTEAATAAGTATMLAMPPEE 350  
 QY 354 LFRADHPFLFYIRKDDI--ILFSGKVS CP 380  
 Db 351 NFNADHPFIFIRHNPSANILFLGRFSSP 379

## RESULT 4

S27383  
 elastase inhibitor - human  
 C:Species: Homo sapiens (man)  
 C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 05-Sep-1997  
 C:Accession: S27383; S65750  
 R:Remold-O'Donnell, E.; Chin, J.; Alberts, M.  
 Proc. Natl. Acad. Sci. U.S.A. 89, 5635-5639, 1992  
 A:Title: Sequence and molecular characterization of human monocyte/neutrophil elastase  
 A:Reference number: S27383; MUID:92302296; PMID:1376927  
 A:Accession: S27383  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 57-65; 97-110; 111-129; 204-213; 216-244; 255-271, 'X', 273-274; 291-301 <PAC>  
 R:Pickard, B.2.; Lee, S.S.; Remold-O'Donnell, E.; Komoriya, A.  
 Blochim. Biophys. Acta 1269, 41-50, 1995  
 A:Title: A serpin from human tumor cells with direct lymphoid immunomodulatory activity:  
 A:Reference number: S65750; MUID:96049524; PMID:7578269  
 A:Accession: S65750  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 57-65; 97-110; 111-129; 204-213; 216-244; 255-271, 'X', 273-274; 291-301 <PAC>  
 C:Genetics:  
 A:Gene: GDB:ELANH2; EI; P12  
 A:Cross-references: GDB:132914; OMIM:130135  
 A:Map position: 6p25-6p24.3  
 C:Superfamily: antithrombin III  
 C:Keywords: serine proteinase inhibitor

Query Match 36.9%; Score 722.5; DB 2; Length 379;  
 Best Local Similarity 38.6%; Pred. No. 1.5e-43;  
 Matches 150; Conservative 74; Mismatches 146; Indels 19; Gaps 4;

QY 1 MASLAANAEEFCFLFREMDDNGNGVFPSSLSLFAALALVRLGADDSLSQIDKLHV 60  
 Db 1 MEOLSSANTRFADLFLALSENNPAGNIFISPFSSISAMAMVFGTGTGNTAAOLSKTFHF 60  
 QY 61 NTASGYCNSSNSGCSOLKRVFSDINASHKVDLSIVNGLFAEKYGGFKHDIIECAEK 120  
 Db 61 NTV-----EVSRSQSLNADINKGASITLKLARLKGKTYNEFLPELVSTOK 110  
 QY 121 LYDAKVERVDFTNHLDETRRNINKWVENETHGKIKNYIGGGISSAVLVNAYVYKRG 180  
 Db 111 TYGADLASVDFOHASEDARTINQWVGOTEGKPELVKGMVDMTKVLVNAIYKGN 170  
 QY 181 MOSAFKSETINCHFKSPKSGKAVAMHOERKFNLSVIDPSMKILELYNG-GIMMYV 239  
 Db 171 WKDFKMEATRDAPFRLNKOTKYKMMYQKKKFPYVYIEDLKCRVLELPYQCKELSMVI 230

QY 240 LLEPN-----DSEIENKLTFLONLMEWTPRRMTSKYVEYFFPOFKIEKNYKQYLRA 293  
 Db 231 LRPDDIEDESTGLEKIEKQLTLEKLRMTKPEMLYLAEVVHLPRFKLEESYDLETSLAR 290  
 QY 294 LGLKDIDESKADLSGIASGRLYISRMHKSYTEVEEGEATAGTSNIVEKQLPOST 353  
 Db 291 LGVODLEFNKRGKADLSGSGARDLFVSKTIHSFVDLEEGTEAATAAGTATMLAMPPEE 350  
 QY 354 LFRADHPFLFYIRKDDI--ILFSGKVS CP 380  
 Db 351 NFNADHPFIFIRHNPSANILFLGRFSSP 379

## RESULT 5

JC7118  
 headpin serine proteinase inhibitor - human  
 C:Species: Homo sapiens (man)  
 C>Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 26-May-2000  
 C:Accession: JC7118  
 R:Spring, P.; Nakashima, T.; Frederick, M.; Henderson, Y.; Clayman, G.  
 Blochim. Biophys. Res. Commun. 264, 289-304, 1999  
 A:Title: Identification and cDNA cloning of headpin, a novel differentially expressed  
 A:Reference number: JC7118; MUID:99458661; PMID:10527881  
 A:Accession: JC7118  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-391 <SPR>  
 A:Cross-references: GB:AF169949; NID:95911368; PIDN:AAD55765.1; PID:95911369  
 C:Genetics:  
 A:Map position: 18q21.3-18q22  
 C:Superfamily: antithrombin III  
 C:Keywords: serine proteinase inhibitor  
 F:355/Inhibitory site: Val (unidentified proteinase) #status predicted

Query Match 36.5%; Score 715.5; DB 2; Length 391;  
 Best Local Similarity 38.0%; Pred. No. 4.8e-43;  
 Matches 152; Conservative 78; Mismatches 141; Indels 29; Gaps 7;

QY 1 MASLAANAEEFCFLFREMDDNGNGVFPSSLSLFAALALVRLGADDSLSQIDKLHV 60  
 Db 1 MDSIGAAYSTRGLFDFEL--KRTNDGNIFPSPVGLTGAIGVLLGTGATASOLEEVFH- 58  
 QY 61 NTASGYCNSSNSGCSOLKRVFSDINASHK-----DYDLSIVNGLPAEK 106  
 Db 59 -----SEKTKSSRIAAEEVEIENEAHQOQKLETSKLTNDYELNTRKLGK 112  
 QY 107 VYGFHKDYIECAEKLVDKVERVDFTNHLDETRRNINKWVENETHGKIKNYIGGGISS 166  
 Db 113 TYLFQKLYLDVEKRYHASLEPVDVNAADSRKINSWESKTEKIKDLFPDGSISS 172  
 QY 167 AVMYLVNAYVFKGKWSAFKSETINCHFKSPKSGKAVAMHOERKFNLSVIDPSMKI 226  
 Db 173 TKLVNVMYVFKGMDPEFKKENTKEKFMWNKSTSKSVOMTOSHSFSFTLDLQAKI 232  
 QY 227 LELKY-NGGINVYLLRPENDLSEIENKLTFLONLMEWTPRRMTSKYVEYFFPOFKIE 282  
 Db 233 LGIYKKNNDLSMEFLLP--NDIDGLEKTIIDKSPKLVEMWSPGMEERKVALHPREVE 291  
 QY 283 KNYEMKQYLRALGLKIDFESKADLSGIASGRLYISRMHKSYTEVEEGEATAGTSN 342  
 Db 292 DSYDLAVLAMGQDAFSEHKADYSGMSSGGLYAKFLHSSFVAVTEGTEAATAAGT 351  
 QY 343 NIVEKQLPOSTLFRADHPFLFYIR--KODIILFSGKVS CP 380  
 Db 352 GFTVTSAPGHENYCHNHPFLFIFIRHNPSANILFLGRFSSP 391

## RESULT 6

DYCH  
 ovalbumin-related y protein - chicken  
 C:Species: Gallus gallus (chicken)  
 C>Date: 17-Dec-1982 #sequence\_revision 17-Dec-1982 #text\_change 18-Jun-1999  
 C:Accession: A01244



```

Query Match      33.9%  Score 664.5;  Db 2:  Length 415;
Best Local Similarity 34.7%  Pred. No. 2,1e-39;
Matches 144;  Conservative 83;  Mismatches 153;  Indels 35;  Gaps 5;

OY 1 MASLAANAACBCEFLFREMDDNCGNCFVSSLSLFLALALVRLGAODDSLQIDKLH 60
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 MEDICVANTLFLALFLFKHLAKASPTQNLFLSPWSISSTMAVMYSGSSTEDQAKYLF 60
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 61 N-----TASGY-----GN-----SSNSOGSLOSQHLRVFSDINASHKD 93
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 NEVGANAVTPMTPENFTSCGFMQIQKSGYPDALIQAAADKIHSSRSLSSAINASTGN 120
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 94 YDLSTVNGCLPAEKYYGFHKDIYIECAEKLYPAKVERVDFTHLDEPTRINIKWVENHFG 153
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 121 YLLESYVKLPEKSKASFREYIRLCOKYYISSEPOAVDFLECAEARKKINSWKTQTKK 180
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 154 IKNVIGEGGSISSAVMYLVNAYVFKGMQSAFTKSETTINCHFSKCSGAIVAMHOERK 213
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

	Query Match	33.8%; Score 662; DB 2;	Length 374;
	Best Local Similarity	37.4%;	Pred. No. 2,7e-39;
	Matches 145;	Conservative 79;	Mismatches 142; Indels 22; Gaps 8;
OY	1	MASIAAANAECFFLRFEMDDNCGNGVAFVSSLSLFALALVRLGAGDSDLQIDKLHH	60
		:     :     :     :     :     :     :     :	
Db	1	MDDICANGTFAISLFKRIIGEEDNRNVFPSPMSISSALAMVFAGAGSTPAOQSOLCL	60
OY	61	NFASGYGNSSNSQGSLQSOLKRVESDIINASHKDVLDSIVNGLFEAKCYGFHKNVIDECAEK	120
		:       :	
Db	61	Y-----KKGDINHGFOSLL-----SEVNRGTOTYLTLTAARLRIGEEKCDLPDPKEKCCK	110
OY	121	LYDAKVERVDFTNHLEDTRRINKWVENETHGIKITNYIGEGISSSAVMVLVAVAAYPKGK	180
		:     :     :     :     :     :     :     :	
Db	111	FYGAELEELSFADTECECRKHINDWVAKEGKI SEYLDAGTVDLPLTKLVLAIVAITYPKG	170
OY	181	WQSAFTSSETINCHEFKSPKCSGKAVALAMHDBERFNLSVIDPSKILIELRY -NGGITMYV	239
		:     :     :     :     :     :     :     :	
Db	171	WNEDFDKRYRGMLETNE--EKTVOIMFR EAKFMQYADEVHTQVLELPEVEBESMVI	229
OY	240	LLEPN--DLSEIENKLTFFONLMETNPTRMTSKYVEVFPQFIKENYAEKQYLRALGLK	297
		:     :     :     :     :     :     :     :	
Db	230	LTPDNLDLVAVERALUYEFKFMATNSEKLTKSXVVQFLPRLRKIEEYDLEPLRRLGMI	289
OY	298	DIEDESAADIASGRLYISRMRKSYILEVTDEGTAAATG---SNIVEQLPOSTL	354
		:     :     :     :     :     :     :	
Db	290	DAPEEAKAADSQGSTENVPLSKVAHACFEVNEDEGETEAAAATAAVRNNSPCSRMEPR--	346
OY	355	FRADHPFLFYIR--KDDILFSGVSCP	380
		:     :     :     :     :     :	
Db	347	FCADHPFLFYIRRKRTKTCILFCGRFSP	374

## RESULT 10

OACH

Ovalbumin [validated] - chicken

C:Species: Gallus gallus (chicken)

C:Date: 31-Dec-1979 #sequence\_revision 30-Jun-1993 #text\_change 15-Sep-2000

C:Accession: A90455; I50402; I50605; A93197; A93827; A90093; A61297; A42793; A01

B:Moore, S.L.C.; Beattie, W.G.; Catterall, J.F.; Dugalczyk, A.; Staden, R.; Brownlee, G.G.

A:Title: Complete nucleotide sequence of the chicken chromosomal ovalbumin gene and its

A:Reference number: A90455; MUID:82069038; PMID:6272839

A:Accession: A90455

A:Molecule type: DNA

A:Residues: 1-386 &lt;MO&gt;

A:Cross-References: EMBL:V00438; NID:963719; PIDN:CAA23716.1; PID:9808974

A:Note: a number of silent polymorphic sites are identified and discussed

A:Note: Thr-188 is also predicted

R:Catterall, J.F.; O'Malley, B.W.; Robertson, M.A.; Staden, R.; Tanaka, Y.; Brownlee, G.

A:Title: Nucleotide sequence homology at 12 intron-exon junctions in the chick ovalbumin

A:Reference number: I50402; MUID:79010682; PMID:692731

A:Accession: I50402

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-386 &lt;CAT&gt;

A:Cross-References: GB:M34352; NID:9212501; PIDN:AAA8998.1; PID:9212503

R:Robertson, M.A.; Staden, R.; Tanaka, Y.; Catterall, J.F.; O'Malley, B.W.; Brownlee, G.

A:Title: Sequence of three introns in the chick ovalbumin gene.

A:Reference number: I50605; MUID:79135070; PMID:423993

A:Accession: I50605

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-4, 'A', 'G', '118', 'F', '120-155 &lt;ROB&gt;

A:Cross-References: EMBL:V00382; NID:963051; PIDN:CAA23681.1; PID:963052

R:McGregor, L.; O'Malley, B.W.; Nisbet, A.D.; Fothergill, J.E.; Givol, D.; Fields, S.

A:Title: Sequence of chicken ovalbumin mRNA.

A:Reference number: A93197; MUID:78199842; PMID:661981

A:Accession: A93197

A:Molecule type: mRNA

A:Residues: 1-386 &lt;MCB&gt;

A:Cross-References: EMBL:V00383; NID:963053

A:Note: a minor component has Asp-312

R:Palmler, R.D.; Gagnon, J.; Walsh, K.A.

Proc. Natl. Acad. Sci. U.S.A. 75, 94-98, 1978

A:Title: Ovalbumin: a secreted protein without a transient hydrophobic leader sequence.

A:Reference number: A93827; MUID:78116057; PMID:272676

A:Accession: A93827

A:Molecule type: protein

A:Residues: 2-33, 'X', '35-36 &lt;PAL&gt;

R:Thompson, E.O.P.; Fisher, W.K.

Aust. J. Biol. Sci. 31, 443-446, 1978

A:Title: A correction and extension of the acetylated amino terminal sequence of ovalbumin.

A:Reference number: A90093; MUID:79186958; PMID:751625

A:Accession: A90093

A:Molecule type: protein

A:Residues: 2-17 &lt;TH1&gt;

R:Thompson, E.O.P.; Fisher, W.K.

Aust. J. Biol. Sci. 31, 433-442, 1978

A:Title: Amino acid sequences containing half-cysteine residues in ovalbumin.

A:Reference number: A90092; MUID:79186957; PMID:751624

A:Accession: A90092

A:Molecule type: protein

A:Residues: 6-17, 30-36, 61-79, 116-124, 367-374, 380-386 &lt;TH2&gt;

R:Tsunasawa, S.; Narita, K.

J. Biochem. 92, 607-613, 1982

A:Title: Micro-identification of amino-terminal acetylamino acids in proteins.

A:Reference number: A61297; MUID:83056735; PMID:6754709

A:Accession: A61297

A:Molecule type: protein

A:Residues: 2-6 &lt;TSU&gt;

R:Takahashi, N.; Hirose, M.

J. Biol. Chem. 267, 11565-11572, 1992

A:Title: Reversible denaturation of disulfide-reduced ovalbumin and its reoxidation

A:Reference number: A42793; MUID:92283876; PMID:1597484

A:Accession: A42793

A:Molecule type: protein

A:Residues: 60-73, 'X', '75-85, 112-119, 'EX', 122-123 &lt;TAK&gt;

R:Stein, P.E.; Leslie, A.G.W.

Submitted to the Brookhaven Protein Data Bank, November 1990

A:Reference number: A50294; PDB:1OVA

A:Contents: annotation; X-ray crystallography, 1.95 angstroms, residues 2-386

A:Status: annotation; X-ray crystallography, 1.95 angstroms, residues 2-386

J. Mol. Biol. 221, 941-959, 1991

A:Title: Crystal structure of unbleached ovalbumin at 1.95 Angstroms resolution.

A:Reference number: A58761; MUID:92046044; PMID:1942038

A:Accession: A58761

A:Contents: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Contents: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

A:Status: annotation; X-ray crystallography, 1.95 angstroms

Query Match	33.6%	Score 658.5;	DB 2;	Length 383;
Best Local Similarity	35.0%;	Pred. No. 4.9e-39;		
Matches 137;	Conservative 96;	Mismatches 139;	Indels 19;	Gaps 7

```

Oy      1 MASLAAANAEPFENFERREDDNOGNGVNFSSJLEPAALVRLGADODSLDKLH 60
Db      1 MGSIAAAMECFEDVFKELKYNHADDNMILSPFALISTAMVFLGKDKDSTRIDINKVHF 60
Oy      61 NTASGCGNSSNSQSG---LOSQLKRVFSDINASHKDYDLSTVNGJFAEKVGFHHDXIE 116
Db      61 DKLPQFGSIEAQGTSVNVHSSLDLNLQITRKQNDAYSFSLASRLAQETIYVPEYIQ 1200
Oy      117 CAEKIYDAKVEYVDTNHL EDTRRINIKKVENETHGKIKNVIGEGCISGAVMYLVNAVY 176
Db      121 CVKELYRGGLSVENFQTADQARGGINAVVESQTIIRNIILQPSVSDQATAVLVNAILA 180
Oy      177 FGKKQSAFPTSETINCHEKSPKSGKAAVAMHNOEKKFNLVSEDEPSMKITLEARY-NGCI 2355
Db      181 FKGLMEKFAFKADDTITPIPRVYTESKPYOMMTIQIGSFVVAASAKSMKILLELPPASGTM 2400
Oy      236 NMVYLLPE--NDLSEIENKLPFQNLMEWTNPRMYTSKYUVEVFPQFIEKNEYEMKQYLRA 2933
Db      241 SMLVLLPDDVSGLEBQLESIFSEKLEWTSSSIEMERKKVYVLLPRKMKMEKYVLTSLMA 3000
Oy      294 LGLKDIPEBSKADLSGIASGRRLYISRMHAKSYIEVTEEGTE--ATAATGSNIVERKOLPO 3511
Db      301 MGIITDLF-SSSANLSCISISVSGSLKISQAVNAHAALINEAGRDVGVSAEAGVDATEE---- 3555
Oy      352 STLPRADHPFLFVIR--KDDILFEGKGVSCP 380
Db      356 ---FRADHPFLCVAHIEFTNALILGRCVSP 383

```

```

RESULT 12
I39184
bomapin - human
N:Alternate names: proteinase inhibitor 10 (PI10)
C:Species: Homo sapiens (man)
C:Date: 01-Mar-1996 #sequence_revision 01-Mar-1996 #text_change 16-Jul-1999
C:Accession: I39184
R:Riewald, M., Schleef, R.
J. Biol. Chem. 270, 26754-26757, 1995
A:Title: Molecular cloning of bomapin (protease inhibitor 10), a novel human serpin that
A:Reference number: I39184; MUID:96070759; PMID:7592909
A:Accession: I39184
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-397 <RES>
A:Cross-references: EMBL:U35459; NID:g1065408; PIDN:AAC50282.1; PID:g1065409
C:Genetics:
A:Gene: GDB:PI10
A:Cross-references: GDB:536283
A:Map position: 14q32.1-14q32.1
C:Superfamily: antithrombin III
C:Keywords: serine proteinase inhibitor

```

Query Match	Similarity	33.6%	Score	658.5	DB	2	Length	397
Best Local	Similarity	34.9%	Pred.	No.5.1e-39				
Matches	Conservative	89	Mismatches	151	Indels	19	Gaps	5
Oy	1	MASLAANAEECFNLFRMEDDNOGNGVNFSSLSFLAALATRLGAODSLSQIDKLHV	60					
		: : : : :   :       : :       : : :   :						
Dd	1	MDSLATSINGALFELSKKLAESAGCKNLFSSWSJSTLTIYYLCAGKTAAQMAYLOF	60					
Oy	61	NPASGYC-----NSSNSQSGLQSLAKVFSDINASHKDYDLSIYNGLPFAEKV	107					
		: : : : :   :   : : :   :   : :     :						
Dd	61	NRDGVKCDPSESEKKRKMEFNLSNEE-IHSPFTLLISEILKPNDYLLKTANAIYGEXT	119					
Oy	108	YGFHKDYTECAENKLDKAVERYDFNHLEDPFRNTINKKVENVETHKIKNVIGEGISSA	167					
		:   :   : : : :   : :   :       :       : :   :						
Dd	120	YAFHNKTYLEDKMTYFGAEPQPVNFEASDAQIRKIDINSVEROTEKIOWLLDDSDVDSTT	179					

[illegible]

```

RESULT 13
A57488
C:proteinase inhibitor Spl3 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 08-Dec-1995 #sequence-revision 08-Dec-1995 #text-change 16-Jul-1999
C:Accession: A57488
R:Sun, J.; Rose, J.B.; Bird, P.
J. Biol. Chem. 270, 16089-16096, 1995
A:Title: Gene structure, chromosomal localization, and expression of the murine homol
A:Reference number: A57488; MUID:95332310; PMID:7608171
A:Accession: A57488
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-378 <SDN>
A:Cross-references: GB:025844; NID:g818902; PIDN:AAA79684.1; PID:g818903
C:Genetics:
A:Gene: Spl3
A:Map position: 13
C:Superfamily: antithrombin III
C:Keywords: serine proteinase inhibitor
E:343/Inhibitory site: Arg (unidentified proteinase) #status predicted
E:343/Inhibitory site: Arg (unidentified proteinase) #status predicted

```

Query Match	33.1%	Score 649	DB 2	Length 378
Best Local Similarity	37.9%	Pred. No. 2,2e-38		
Matches	146	Conservative	77	Mismatches 150; Indels 12; Gaps 7
QY	1	MASLAAANAECFCNLFREMDNDGNGNVFFSSLSFLPALALVRLGAQDDLSQIDKLHV	60	
DB	1	MDPLQANGTALNALTKLILGEDSSK-NYFLSPMSISISALAMFAMKAGTTASOMAOALAL	59	
QY	61	NTASGYSNNSQSGLOSOLKRVPSDINASHHDQDLSYNGCFEAKYGVGFHNDYIECAEK	120	
DB	60	DKCSNGGGDVHQ-GFOSLL---TEVNKTGTQYLLRTANRLFGKTCDDLASFQDSCLK	114	
QY	121	LYDAKVERVEFTNHLDEPTRNINKWVENETNKKIKNVIGEGGISSAAVVLVNAVYFNGK	180	
DB	115	FYEALDELDELQGAATEESRQHNTWVAKKTEDEKIEVLSPTGVNSDTSLYLVNALYFNGN	174	
QY	181	WOSAFTKSEETINCHEKSPKSCGAKAAMMHQEFKFLSYEDPSMKILERY-NGGINMY	239	
DB	175	WEKQNKMKHTREMPKVKVSKNEBKPOAMFKKSTFMITYIGELFTFTLILLPYVSSBLMI	234	
QY	240	LLPEN--DLSELENKLTQFONLEMTNPNRMTKSYVEVFPOPKIKENYEMKOYLALGK	297	
DB	235	MLPDRHVELSYVEKEVLYEKPLETRLRDKMDEBEVPLPKKLLEENTNMNALKKLGMT	294	
QY	298	DIFDESKADLSGIASGRGLYISRMHKSYLEVEEGTEATATAGSNIYEKOLPOSTLPR	357	
DB	295	DAFG-GRADFESGMSKQGLFLSKVYHKKFVEVNEGTEAATAATAGMYTCRMFRPRECA	353	
QY	358	DHPLEFLVIR--KDDILFSGKVS	380	
DB	354	DHPLEFLTHVKTNGILFCGRFSS	378	

## RESULT 14

A48681  
Placental thrombin inhibitor - human  
N:Alternate names: cytoplasmic antiproteinase; intracellular serine proteinase inhibitor  
C:Species: Homo sapiens (man)  
D:Date: 07-Apr-1994 #sequence revision 07-Jul-1995 #text change 18-Jun-1999  
C:Accession: A48681; A54352; A46672; B46672; C46672; S35750  
R:Coughlin, P.; Sun, J.; Cerruti, L.; Salem, H.H.; Bird, P.  
Proc. Natl. Acad. Sci. U.S.A. 90, 9417-9421, 1993  
A>Title: Cloning and molecular characterization of a human intracellular serine proteinase  
A:Reference number: A48681; MUID:94022386; PMID:8415716  
A:Accession: A48681  
A:Molecule type: mRNA  
A:Residues: 1-376 <COO>  
A:Cross-references: GB:222658; NID:g297411; PIDN:CAA80373.1; PID:g297412  
A:Experimental source: Placenta  
A>Note: Authors translated the codon CAA for residue 198 as Gly  
R:Morgenstern, K.A.; Sprecher, C.; Holtz, L.; Foster, D.; Grant, F.J.; Ching, A.; Kistler  
Biochemistry 33, 3432-3441, 1994  
A>Title: Complementary DNA cloning and kinetic characterization of a novel intracellular  
A:Reference number: A54352; MUID:94183847; PMID:8136380  
A:Accession: A54352  
A:Molecule type: mRNA  
A:Residues: 1-174,'E','S',363-376 <MOR>  
A:Cross-references: GB:S69272; NID:g546087; PIDN:AAB30320.1; PID:g546088  
A:Experimental source: Placenta  
A>Note: sequence modified after extraction from NCBI backbone  
C:Genetics:  
A:Gene: GDB:PT6  
A:Cross-references: GDB:252025; OMIM:173321  
A:Map position: 6p25-6p24.3  
C:Superfamily: antithrombin III  
C:Keywords: blocked amino end; cytosol; serine proteinase inhibitor  
F:341/Inhibitory site: Arg (thrombin) #status predicted

Query Match 33.1%, Score 648; DB 1; Length 376;  
Best Local Similarity 39.1%, Pred. No. 2.6e-38;  
Matches 151; Conservative 74; Mismatches 145; Indels 16; Gaps 8;

OY 1 MASLAANAEECPFLPREM--DDNGGNGVPEPSSLSLPAALATVLRGADDSLODKLH 59  
Db 1 MDVLAENAGTFLNLTLTKGN--SKNVFPSPMSICALAMVMYGAKGNIAQMAAILIS 58  
OY 60 VNTASGYGNSSNSOSGLQLKRFPSDINASHKYDIASIYGVLEAEVYGFHKIYICAE 119  
Db 59 FNKSGGGG---DHQGQSLL---TEVNKTGTGYLLRVANRLPFGEKSDFELSFRRSCQ 111  
OY 120 KLIDAKYERVDFTNHLEDTRRNINKWVENETHGIKNVIGEGISSAVLVNAVYFKG 179  
Db 112 KFYQAEMBEDLFISAVERSKRHINTWVAEKTEGIAELLSPGSVDPLRLVLNAVYFRG 171  
OY 180 KWGAAPFKSEIRINHFFSPKCSGKAVAMMHDEKFNLSVIEDPSMKLTLELRYN--GINMY 238  
Db 172 NMDCQPKRENTERRFLPVSKNEKPVOVMRKQSTFFKTYTGELFTQTLLVPYGEKLMI 231  
OY 239 VLIDE--NDLSEINLKTLFONLMETNPBRRTSKYVEVFPPQFIKENYEMKOYLRAIGL 296  
Db 232 IMPDETDTDLRTVKELTFEYFWTRLDMADEDEVEVSLPRKLESISTYMESYLRNLGM 291  
OY 297 KDIFESKADLSGIASGRRLYSIPMMHRSYLEVEEBCTEATAAGSNIVIKOLPOSTLFR 356  
Db 292 TDARELCKADPSGH--SQTDLSLSKVHSFYEVAEBESTFAAAAATAAIIIMMRCAAFVPRFC 350  
OY 357 ADHPFLVI--RKDDITLFGSKVSCP 380  
||||||| | | : ||| | | : |

Db 351 ADHPLEFIIQHRKTNGILFCGRFSSP 376

RESULT 15

S20047  
plasminogen activator inhibitor 2 - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 07-Apr-1994 #sequence\_revision 07-Apr-1994 #text\_change 16-Jul-1999  
C:Accession: S20047; S09616  
R:Belin, D.  
submitted to the EMBL Data Library, September 1989  
A:Reference number: S20047  
A:Accession: S20047  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-415 <BEL>  
A:Cross-references: EMBL:X16490; NID:953589; PIDN:CA34507.1; PID:953590  
R:Belin, D.; Wohlwend, A.; Schleuning, W.D.; Kruthof, E.K.O.; Vassalli, J.D.  
EMBO J. 8, 3287-3294, 1989  
A:Title: Facultative polypeptide translocation allows a single mRNA to encode the sec  
A:Reference number: S09615; MUID:9005920; PMID:2583099  
A:Accession: S09616  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-27 <BEL2>  
A:Cross-references: EMBL:X16490  
C:Superfamily: antithrombin III

Query Match:	32.2%	Score 630.5:	DB 2:	Length 415:
Best Local Similarity	34.4%	Pred. No. 5.2e-37:		
Matches 145:	Conservative 75:	Mismatches 155:	Indels 49:	Gaps 8

  

QY	1	MASLAANA	EFCNLFREMDNOGNGNVFFSSLSLPAALALVRLGAODDLSIQDKLLHV	60
Db	1	MELISAANTM	FALNLLKQIEKSNSTONIFISPWSSISSTLAVLLGAGNTEQOMAKVLQF	60
QY	61	NFASGYG----	SSN-----SSG--LQSLQKVPESDINASHKD	93
Db	61	NELGSGITTRN	PENFSGCDPAQOQOKENYBPAILQAOAGKTHSASSLSSTINTPGCD	120
QY	94	YDLISVNGL	FAEKVYGFHKDYIECAEKLVDAKERVDFTNHLEDTRRNINKWENETHGK	153
Db	121	YLLESANKL	FEKESAFKEEYIQLSKYYSTPEPENVDFLECAEBAREKINSWATQTKGE	180
QY	154	IKKIVIEGG	ISSASVAVLVNAVYFPGKMOASFTKSET-----INCHPKSCGSKAVAM	207
Db	181	IPMLDEG	SVDETKMVLVNAVYFPGKMKTPPEKKLNGLYFFRNHSHESP-----VQM	234
QY	208	MHOERFNL	SVIEDPSMKTLELRNGINMYVLP-----NDLSEIENKLFQONLMEW	261
Db	235	MFLHALNT	IGYIKDKTQITLPHNGISMILLTLPDEIEDASTGLELSEINFRANKW	294
QY	262	TNFRRTSK	SVYVEFPFPEKTYENMKOYLRLGLKDFIDASKDLGSIAGSGRLYFSRM	321
Db	295	ISKQITL	DEDDVVYIIPKRLAOSIYELKSILOSQMGEDAFNGKKNFSGMSBRNDLPFSEV	354
QY	322	MKRSYIEV	TEGETAATGSGNIYERKODPOSTFLRADHPLFLVIIRKDI--ILFSGKVS	378
Db	355	FHOASVDY	TEEGTVAAGTGAVMTGRTHGGPQFVADHPFLFI-MDKITHITLIFVGRFS	413
QY	379	CP	380	
Db	414	SP	415	

Search completed: July 11, 2003, 12:00:45  
Job time : 77.732 secs

357 ADHPFLFVI--RKDDILFSGKVS CP 380



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:43:41 ; Search time 36.341 Seconds  
(without alignments)  
433.698 Million cell updates/sec

Title: US-10-091-442-34  
Perfect score: 1959  
Sequence: 1 MASLAANAEFCNLFREMD.....FLFVIRKDDILFSGKVSQP 380

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues  
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query %	Match length	ID	Description
1	1959	100.0	380	1 SPB7_HUMAN	O75635 homo sapien
2	819	41.8	392	1 SB11_HUMAN	O96p15 homo sapien
3	763.5	39.0	405	1 SB12_HUMAN	O96p63 homo sapien
4	752	38.4	390	1 SCC1_HUMAN	P48598 homo sapien
5	748	38.2	390	1 SCC2_HUMAN	P48594 homo sapien
6	739.5	37.7	379	1 ILEU_HORSE	P05619 equus caball
7	722.5	36.9	379	1 ILEU_HUMAN	P30740 homo sapien
8	714.5	36.5	391	1 SB13_HUMAN	O9u1v8 homo sapien
9	711	36.3	388	1 OVAY_CHICK	P01014 gallus gall
10	702	35.8	378	1 PT16_BOVIN	O02739 bos taurus
11	696	35.5	378	1 ILEU_PIG	P80229 sus scrofa
12	664.5	33.9	415	1 PA12_HUMAN	P05120 homo sapien
13	662	33.8	374	1 SPB8_HUMAN	P50452 homo sapien
14	658.5	33.6	397	1 SB10_HUMAN	P48595 homo sapien
15	654	33.4	385	1 OVAL_CHICK	P01012 gallus gall
16	653.5	33.4	382	1 OVAL_COTJA	P19104 coturnix co
17	649	33.1	376	1 PT16_MOUSE	O60854 mus musculu
18	644	32.9	376	1 PT16_HUMAN	P35227 homo sapien
19	630.5	32.2	415	1 PA12_MOUSE	P12338 mus musculu
20	623	31.8	376	1 SPB9_HUMAN	P50453 homo sapien
21	622	31.8	416	1 PA12_RAT	P29534 rattus norv
22	605.5	30.9	375	1 MASP_RAT	P70124 rattus norv
23	600.5	30.7	375	1 MASP_MOUSE	P32261 mus musculu
24	592	30.2	465	1 ANT3_MOUSE	P32261 mus musculu
25	590.5	30.1	375	1 MASP_HUMAN	P36932 homo sapien
26	567	28.9	433	1 ANT3_BOVIN	P41361 bos taurus
27	566	28.4	464	1 ANT3_HUMAN	P10008 homo sapien
28	557	28.4	465	1 ANT3_SHEEP	P32262 ovis aries
29	503.5	25.7	410	1 NEUS_RAT	O91142 rattus norv
30	498.5	25.4	410	1 NEUS_HUMAN	O99574 homo sapien
31	492	25.1	405	1 SP12_MOUSE	O91k88 mus musculu
32	484.5	24.7	410	1 NEUS_MOUSE	O35664 mus musculu
33	478.5	24.4	417	1 KBP_MOUSE	P29621 mus musculu

## ALIGNMENTS

RESULT 1	SPB7_HUMAN	STANDARD:	PRT:	380 AA.
AC	O75635			
DT	15-JUN-2002 (Rel. 41, Created)			
DT	15-JUN-2002 (Rel. 41, Last sequence update)			
DE	Megsin (TPP55) (Serpin B7).			
GN	SERPINF7			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=97326116; PubMed=9182567;			
RA	Tsujimoto M., Tsunoka N., Ishida N., Kurihara T., Iwasa F.,			
RA	Yamashiro K., Rogi T., Kodama S., Katsuragi N., Adachi M.,			
RA	Katayama T., Nakao M., Yamachi K., Hashino J., Haryama M., Miura K.,			
RA	Nakanishi T., Nakazato H., Teramura M., Mizoguchi H., Yamaguchi N.,			
RT	"Purification, cDNA cloning, and characterization of a new serpin with			
RT	megakaryocyte maturation activity."			
RL	J. Biol. Chem. 272:15373-15380(1997).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Mesangial cells;			
RX	MEDLINE=98376492; PubMed=9710452;			
RA	Miyata T., Nangaku M., Suzuki D., Inagi R., Uragami K., Sakai H.,			
RA	Okubo K., Kurokawa K.;			
RT	"A mesangium-predominant gene, meslin, is a new serpin upregulated in			
RT	IGA nephropathy."			
RT	J. Clin. Invest. 102:828-836(1998).			
CC	- FUNCTION: Might influence the maturation of megakaryocytes via			
CC	proteases. Might influence the maturation of Lys-specific			
CC	its action as a serpin.			
CC	- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).			
CC	- TISSUE SPECIFICITY: Predominantly expressed in mesangial cells.			
CC	- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPINF SUBFAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL: D88575; BAA31232.1; -			O90935 gallus gall
DR	EMBL: AF027866; AAC64506.1; -			P01011 homo sapien
DR	HSSP: P05619; IHLF.			O75830 mus musculu
DR	Genew: HGNC:13902; SERPINB7.			O03044 didelphis m
DR	MTM: 603357; -			P07739 mus musculu
DR	InterPro: IPR000215; Serpin.			O60366 apodemus sy
DR	Pfam: PF00079; serpin; 1.			O54758 tamias sibi
DR	SMART: SM00093; SERPIN; 1.			P09006 rattus norv
DR	PROSITE: PS00284; SERPIN; 1.			P70458 mus musculu
				P22325 cavia porce
				P05154 homo sapien
				O54757 tamias sibi

KW Serpin; Serine protease inhibitor.  
 FT ACT\_SITE 347 348 REACTIVE BOND (BY SIMILARITY).  
 SQ SEQUENCE 380 AA; 42904 MW; 9A2CDB6C63CF605 CRC64;

Query Match 100.0%; Score 1959; DB 1; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 1,2e-129;  
 Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASLAANAEFCNLFREDDNDGNGNVEFSSLSLFAALALVRLGAQDLSIQIDKLHAY 60  
 DB 1 MASLAANAEFCNLFREDDNDGNGNVEFSSLSLFAALALVRLGAQDLSIQIDKLHAY 60  
 QY 61 NTASGCGNSNSGSGLOSLKRVFSDINASHKDYDLSYNGLFAEKYVGFHKDYTECAEK 120  
 DB 61 NTASGCGNSNSGSGLOSLKRVFSDINASHKDYDLSYNGLFAEKYVGFHKDYTECAEK 120  
 QY 121 LVDKAEVDFPNHLEDTFRNINKWVENETHGKIKRVIGEGGSSSAVAVLVNAVYFKK 180  
 DB 121 LVDKAEVDFPNHLEDTFRNINKWVENETHGKIKRVIGEGGSSSAVAVLVNAVYFKK 180  
 QY 181 WQSAPFSEITNCHFKSPKCGKAVAMHGERKFNLSYEDPSMKILELRNGIMAYVL 240  
 DB 181 WQSAPFSEITNCHFKSPKCGKAVAMHGERKFNLSYEDPSMKILELRNGIMAYVL 240  
 QY 241 LPENDLSEIENKLTFONLMEWTNPRMTSKYVEVFPQFKIEKNEMKOYLRLALGKIDF 300  
 DB 241 LPENDLSEIENKLTFONLMEWTNPRMTSKYVEVFPQFKIEKNEMKOYLRLALGKIDF 300  
 QY 301 DESKADLSGIASGGRLYISRMHKSLEYTEEGTGTATGSIYVKOLPOSTLFRADHP 360  
 DB 301 DESKADLSGIASGGRLYISRMHKSLEYTEEGTGTATGSIYVKOLPOSTLFRADHP 360  
 QY 361 FLEVIKRDIIILFSGKVSCP 380  
 DB 361 FLEVIKRDIIILFSGKVSCP 380

## RESULT 2

ID SB11\_HUMAN STANDARD; PRT; 392 AA.  
 AC Q96P15; Q96P14; Q96P13;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Serpin B11.  
 GN SERPINB11.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND VARIANTS THR-148 AND THR-181.  
 RA Silverman G.A., Cataltepe S.;  
 RT "SERPINB11 is a novel ov-serpin with at least 3 alleles."  
 RT Submitted (SEP-2001) to the EMBL/Genbank/DBD databases.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: AF419953; AAL16056.1; -  
 DR EMBL: AF419954; AAL16057.1; -  
 DR EMBL: AF419955; AAL16058.1; -  
 DR GenBank: HGNC:14221; SERPINB11.  
 DR InterPro: IPR000215; Serpin.  
 DR Pfam: PF00079; Serpin; 1.  
 DR PROSITE: PS00284; SERPIN, 1.

KW Serpin; Serine protease inhibitor; Polymorphism.  
 FT ACT\_SITE 357 358 REACTIVE BOND (BY SIMILARITY).  
 FT VARIANT 148 148 M -> T (IN ALLELE B).  
 FT VARIANT 181 181 /FTID-VAR\_012472  
 FT VARIANT 181 181 I -> T (IN ALLELE B AND ALLELE C).  
 FT /FTID-VAR\_012473.  
 SQ SEQUENCE 392 AA; 44098 MW; 906F6D412BC0756 CRC64;

Query Match 41.8%; Score 819; DB 1; Length 392;  
 Best Local Similarity 43.1%; Pred. No. 4,1e-50;  
 Matches 169; Conservative 80; Mismatches 131; Indels 12; Gaps 6;

QY 1 MASLAANAEFCNLFREDDNDGNGNVEFSSLSLFAALALVRLGAQDLSIQIDKLHAY 60  
 DB 1 MASLAANAEFCNLFREDDNDGNGNVEFSSLSLFAALALVRLGAQDLSIQIDKLHAY 60  
 QY 61 NTASGCGNSNSGSGLOSLKRVFSDINASHKDYDLSYNGLFAEKYVGFHKD 113  
 DB 61 NTASGCGNSNSGSGLOSLKRVFSDINASHKDYDLSYNGLFAEKYVGFHKD 113  
 QY 121 LVDKAEVDFPNHLEDTFRNINKWVENETHGKIKRVIGEGGSSSAVAVLVNAVYFKK 180  
 DB 121 LVDKAEVDFPNHLEDTFRNINKWVENETHGKIKRVIGEGGSSSAVAVLVNAVYFKK 180  
 QY 181 WQSAPFSEITNCHFKSPKCGKAVAMHGERKFNLSYEDPSMKILELRNGIMAYVL 240  
 DB 181 WQSAPFSEITNCHFKSPKCGKAVAMHGERKFNLSYEDPSMKILELRNGIMAYVL 240  
 QY 241 LPENDLSEIENKLTFONLMEWTNPRMTSKYVEVFPQFKIEKNEMKOYLRLALGKIDF 300  
 DB 241 LPENDLSEIENKLTFONLMEWTNPRMTSKYVEVFPQFKIEKNEMKOYLRLALGKIDF 300  
 QY 301 DESKADLSGIASGGRLYISRMHKSLEYTEEGTGTATGSIYVKOLPOSTLFRADHP 360  
 DB 301 DESKADLSGIASGGRLYISRMHKSLEYTEEGTGTATGSIYVKOLPOSTLFRADHP 360  
 QY 361 FLEVIKRDIIILFSGKVSCP 380  
 DB 361 FLEVIKRDIIILFSGKVSCP 380

## RESULT 3

ID SB12\_HUMAN STANDARD; PRT; 405 AA.  
 AC Q96P63;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Serpin B12.  
 GN SERPINB12.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RA MEDLINE-21623618; PubMed-11604408;  
 RA Askew Y.S., Pak S.C., Luke C.J., Askew D.J., Cataltepe S., Mills D.R.,  
 RA Kato H., Lebowitz J., Dewar K., Birren B., Silverman G.A.;  
 RT "SERPINB12 is a novel member of the human ov-serpin family that is  
 RT widely expressed and inhibits trypsin-like serine proteinases."  
 RT J. Biol. Chem. 276:49320-49330(2001).  
 CC -1- FUNCTION: Inhibits trypsin and plasmin, but not thrombin,  
 CC coagulation factor Xa, or urokinase-type plasminogen activator.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- TISSUE SPECIFICITY: Expressed in many tissues, including brain,  
 CC bone marrow, lymph node, heart, lung, liver, pancreas, testis,  
 CC ovary, and intestine.  
 CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AF11191; AL05571.1; -  
 DR Genew: HGNC:14220; SERPIN12.  
 DR InterPro: IPR000215; Serpin.  
 DR Pfam: PF00079; serpin.1.  
 DR Prosite: PS00284; SERPIN.1.  
 KM Serpin: Serine protease inhibitor.  
 FT ACT\_SITE 370 371 REACTIVE BOND (BY SIMILARITY).  
 SQ SEQUENCE 405 AA; 46276 MW; FFE12DAC9B7F3DPA CRC64;

Query Match 39.0%; Score 763.5; DB 1; Length 405;  
 Best Local Similarity 39.3%; Pred. No. 3.2e-46;  
 Matches 155; Conservative 87; Mismatches 114; Indels 25; Gaps 5;

QY 1 MASLAANAEEFCNLFREMDNOGNVFFSSLSLFAALALVRLGADDSLSQIDKLHV 60  
 DB 1 MDSLVTANTKFCFDLPFGKIDRKHKIFPSPLSLAALGMVRLGARSDSAHOIDEVLHF 60

QY 61 NTAS-----GYNSSNSQGLQS-OLKRVPSDINASHKDYDLTVNGLEFA 104  
 DB 61 NEFSQNSKEPDPCLKSNKOKAGSLNNSGVLSCYFQOLLKIDRIKTDTLSIANRLYG 120

QY 105 EKVYGFHKDYIECAEKLYDAKVERVDFTNLEDRRNINRWENETGKIKNVGEGIS 164  
 DB 121 EQEFPICQYLDVGYQYHTTIESVDQKPKERSKQENRWECOSQKIKELFSKALIN 180

QY 165 SSAVMVLNVNVYFKGQKQSAFTSEKTSINCHFKSPKCGKAVAMHQRKKNLSYIEDPSM 224  
 DB 181 AEVLVVLNVNVYFKAKETFEHDENYDAPFCLANENKSKVMKQGLRIGTIEVKA 240

QY 225 KILELRK-NGGINMYLLPND-----LSEIENKLTFOULMEYTNPRRTSKYVEVFP 277  
 DB 241 QILEMRYTKGLKSLFVLLPSSHDKNLKGLLEELERKIYEKKVVAASSSESVLSPF 300

QY 278 OFIKERKYEKQYLRALGLDIFDESKADLSGASGRLYISRMMHKSIEVTEGTEAT 337  
 DB 301 RFLLEDSTDLNLSLQDGLDITDIEDRADLGLSPSPNLKSLTIHRTFVEVDENGTOAA 360

QY 338 AATGSNIVERKQLPOSTLFRADHPFLFYIR--KDDILFSGKVSQP 380  
 DB 361 AATGAVAVSERLSRWSWFEFNNHPLFFIRHNKQTQILFYGRVSCP 405

RESULT 4  
 SCCL\_HUMAN  
 ID SCCL\_HUMAN STANDARD; PRT; 390 AA.  
 AC P29508: 096J21;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Squamous cell carcinoma antigen 1 (SCCA-1) (Protein T4-A).  
 GN SERPIN3 OR SCCA1 OR SCCA.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 ON NCBI\_TaxID:9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.  
 RX MEDLINE-92068241; Pubmed-1958219.  
 RA Suminami Y., Kishi F., Sekiguchi K., Kato H.;  
 RT "Squamous cell carcinoma antigen is a new member of the serine  
 RT protease inhibitors.";  
 RL Biochem. Biophys. Res. Commun. 181:51-58(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-95241462; Pubmed-7724531;  
 RA Schneider S.S., Schick C., Fish K.E., Miller E., Pena J.C.,  
 RA Treter S.D., Hui S.M., Silverman G.A.;  
 RT "A serine proteinase inhibitor locus at 18q21.3 contains a tandem

RT duplication of the human squamous cell carcinoma antigen gene.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:3147-3151(1995).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lung;  
 RA Strausberg R.;  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: MAY ACT AS A PROTEASE INHIBITOR TO MODULATE THE HOST  
 CC IMMUNE RESPONSE AGAINST TUMOR CELLS.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- TISSUE SPECIFICITY: SQUAMOUS CELLS.  
 CC -1- DEVELOPMENTAL STAGE: ITS EXPRESSION IS CLOSELY RELATED TO CELLULAR  
 CC DIFFERENTIATION IN BOTH NORMAL AND MALIGNANT SQUAMOUS CELLS.  
 CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL: S66896; AAB20405.1; -  
 DR EMBL: U19556; AAA97552.1; -  
 DR EMBL: U19568; AAA86317.1; -  
 DR EMBL: U19559; AAA86317.1; JOINED.  
 DR EMBL: U19560; AAA86317.1; JOINED.  
 DR EMBL: U19565; AAA86317.1; JOINED.  
 DR EMBL: U19567; AAA86317.1; JOINED.  
 DR EMBL: U19562; AAA86316.1; JOINED.  
 DR EMBL: U19560; AAA86316.1; JOINED.  
 DR EMBL: U19560; AAA86316.1; JOINED.  
 DR EMBL: BC005224; AAH05224.1; -  
 DR PIR: J10966; J10966.  
 DR HSSP: P01008; IATH.  
 DR Genew: HGNC:10569; SERPINB3.  
 DR MIM: 600517; -  
 DR InterPro: IPR000215; Serpin.  
 DR Pfam: PF00079; serpin.1.  
 DR SMART: SM00093; SERPIN.1.  
 DR Prosite: PS00284; SERPIN.1.  
 KM Serpin: Serine protease inhibitor.  
 FT ACT\_SITE 354 355 REACTIVE BOND.  
 FT CONFLICT 357 357 T -> A (IN REF. 1).  
 SQ SEQUENCE 390 AA; 44564 MW; B5F27F986C752CFA CRC64;

Query Match 38.4%; Score 752; DB 1; Length 390;  
 Best Local Similarity 41.5%; Pred. No. 1.9e-45;  
 Matches 165; Conservative 81; Mismatches 126; Indels 26; Gaps 10;

QY 1 MASLAANAEEFCNLFREMDNOGNVFFSSLSLFAALALVRLGADDSLSQIDKLHV 60  
 DB 1 MNSLSEANTKFMFLDFOQFRKSEN-NIFSPISITSALGAVLLGAKDNTAQKRYLHF 59

QY 61 -----NTASGYNSSNSQGL-LOSQLRVPSDINASHKQVDLSVNLFLFAKYGKQY 114  
 DB 60 DQVTEITGKAATYHVDSNVHNOFQKLLTEFNKSYDAELKLANLFEKTYLFLQY 119

QY 115 IECAEKLYDAKVERVDFTNLEDRRNINRWENETGKIKNVIGEGISSAVMYLVNA 174  
 DB 120 LDAIKRTYQTSVEVDAPNAPPEERKKNINSVESQTEKIKNLIPENISNTTLVYNA 179

QY 175 VYFKGQKQSAFTSEKTSINCHFKSPKCGKAVAMHQRKKNLSYIEDPSKILELRNG- 233  
 DB 180 IYFGQWEKKFNKEDTKEEFKFWPKNTYKSIQWRYQTSFHFASLSDVQAKYLEIPYK 239

QY 234 GINMYVLLP--NDLSIENKLTFOULMEYTNPRRTSKYVEVFPQFKIEKYEKQY 291  
 DB 240 DLSIVVLLPNDLQGLDGLKLEKLTAEKLEWTSIQNMKRETRVDLILPFEKVEESYDLKDTL 299

QY 292 RALGLKIDFDESKADLSGASGRLYISRMMHKSIEVTEGTEAT-----GSNIVE 346

```

Db      300 RTMGWNIJFN-GDADLSCGTMGSRGLSCVLAHAFVEVEEGEAAAATAVVGFGSS--- 355
QY      347 KOLPOST--LFRADHPFLFVIR--KDDILFSGKVCSP 380
      356 ---PTSTNEFHCNHFFLFPIRONTKNSILFGRFSSP 390

RESULT 5
SCC2_HUMAN
ID      SCC2_HUMAN          STANDARD:          PRT:          390 AA.
AC      P48594;
DT      01-FEB-1996 (Rel. 33, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DE      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Squamous cell carcinoma antigen 2 (SCCA-2) (leuplin).
GN      SERPINB4 OR SCCA2.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
      [1] NCBI_TaxID=9606;
RP      SEQUENCE FROM N.A.
RX      MEDLINE=95241462; PubMed=7724531;
RA      Schneider S.S., Schick C., Fish K.E., Miller E., Pena J.C.,
RA      Treter S.D., Hui S.M., Silverman G.A.;
RT      "A serine proteinase inhibitor locus at 18q21.3 contains a tandem
RT      duplication of the human squamous cell carcinoma antigen gene.";
RL      Proc. Natl. Acad. Sci. U.S.A. 92:3147-3151(1995).
      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=96013887; PubMed=7589435;
RA      Barnes R.C., Morrell D.M.;
RT      "Identification of a novel human serpin gene: cloning sequencing and
RT      expression of leupin.";
RL      FEBS Lett. 373:61-65(1995).
      [3]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=21167379; PubMed=11267667;
RA      Hamada K., Shitomiya H., Asano Y., Kihana T., Iwamoto M., Hanakawa Y.,
RA      Hashimoto K., Hirose S., Kyo S., Ito M.;
RT      "Molecular cloning of human squamous cell carcinoma antigen 1 gene and
RT      characterization of its promoter.";
RL      Biochim. Biophys. Acta 1518:124-131(2001).
      [4]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Bone marrow;
RA      Strausberg R.;
RL      Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
CC      -1- FUNCTION: MAY ACT AS A PROTEASE INHIBITOR TO MODULATE THE HOST
CC      IMMUNE RESPONSE AGAINST TUMOR CELLS.
CC      -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC      -1- TISSUE SPECIFICITY: SQUAMOUS CELLS.
CC      -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; U19576; AAA92602.1; -
DR      EMBL; U19570; AAA92602.1; JOINED.
DR      EMBL; U19571; AAA92602.1; JOINED.
DR      EMBL; U19572; AAA92602.1; JOINED.
DR      EMBL; U19574; AAA92602.1; JOINED.
DR      EMBL; U19575; AAA92602.1; JOINED.
DR      EMBL; U19557; AAA92602.1; JOINED.
DR      EMBL; X89015; CAA61420.1; -
DR      EMBL; AB035089; BAB21525.1; -
DR      EMBL; BC017401; AAH17401.1; -

```

```

DR      HSSP; P01008; 1AT8.
DR      Genew; HGNC:10570; SERPINB4.
DR      MIM; 600518; -
DR      InterPro; IPR000215; Serpin.
DR      Pfam; PF00079; serpin. 1.
DR      SMART; SM00093; SERPIN; 1.
DR      PROSITE; PS00284; SERPIN; 1.
KW      Serpin; Serine protease inhibitor.
FT      ACT_SITE 354 355
SQ      SEQUENCE 390 AA; 44854 MW; 04E213CD892587D5 CRC64;

Query Match      38.2%; Score 748; DB 1; Length 390;
Best Local Similarity 40.8%; Pred. No. 3,6e-45;
Matches 160; Conservative 84; Mismatches 134; Indels 14; Gaps 8;

QY      1 MASIAAANPECFNIFREDDNOGNGNVFESSLSFAALATVRLGAODDSLODKLHV 60
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB      1 MNSLSEANRKFMDLEQQRKSKEN-NIFSPSTSLGAVLLAKDKNTAQOISVLHF 59
QY      61 -----NTASGYGNSSNSQSG-LSQQLRVFSDINASHKDYLSIVNGLFAEKYGFHKDY 114
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB      60 DQVTEHTTEKATYHYHDSRGVNHQFOKLTLEFNKSTDAVELKANKLGEKTYQELQY 119
QY      115 ITCAEKLYDAKYERVDFTNHLIEDTRRNINKWVENETHGKIKVIGBGISSAVAVLVNA 174
      ::|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB      120 LPAIKKFYQTSVESTDFANAPESRKRKINSWESQTEKIKMLFPDGTIGNDTTLVLVNA 179
QY      175 VYFKGKWSAFKSETINCHFKSPKCSKAVAMHMOERFNLVIEDPSMKLTLELRNG- 233
      :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB      180 IIFKGWENKFKKENTKEKFWPNKNTYKSVMMQYNSFNALLEDVQAKLEIPYKCK 239
QY      234 GIMVYVLPPE--NDLSELENKLTFLQNLMEVTPRRKTSKYVEVFPQFIENKYMOKYL 291
      ::|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB      240 DLSMIVLPENEDIGLQKLEKLTAEKLMWMTSLQNMRETCVDLHLPREFMESSYDLKDTL 299
QY      292 RALGKIDFDESKADLSIAGSGRLYISMMMKSTIEVTEETATGSIYVEKQDPQ 351
      |::|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB      300 RTMGWNIJFN-GDADLSCGTMGSHGLSVSLKAKAVEVEEGEAAAATAVVGVELSPS 358
QY      352 ST-LFRADHPFLFVIR--KDDILFSGKVCSP 380
      :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB      359 TNEFHCNHFFLFPIRONTKNSILFGRFSSP 390

RESULT 6
ILEU_HORSE
ID      ILEU_HORSE          STANDARD:          PRT:          379 AA.
AC      P05619;
DT      01-NOV-1988 (Rel. 09, Created)
DT      01-FEB-1994 (Rel. 28, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Leucocyte elastase inhibitor (LEI).
GN      SERPINB1 OR ELANH2.
OS      Equus caballus (Horse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX      Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
      [1] NCBI_TaxID=9796;
RP      SEQUENCE FROM N.A.
RX      MEDLINE=93319507; PubMed=7687128;
RA      Kordula T., Dublin A., Schooltink H., Koj A., Heinrich P.C.,
RA      Rose-John S.;
RT      "Molecular cloning and expression of an intracellular serpin: an
RT      elastase inhibitor from horse leucocytes.";
RL      Biochem. J. 293:187-193(1993).
      [2]
RP      SEQUENCE.
RC      TISSUE=Leukocyte;
RX      MEDLINE=92202200; PubMed=1551869;
RA      Dublin A., Travis J., Enghild J.J., Potempa J.;
RT      "Equine leucocyte elastase inhibitor. Primary structure and
RT      identification as a thymosin-binding protein.";
RL      J. Biol. Chem. 267:6576-6583(1992).

```

RN [3]  
 RP SEQUENCE OF 343-362.  
 RX MEDLINE=88213423; PubMed=3366785;  
 RA Potempa J., Dubin A., Watorek W., Travis J.;  
 RT "An elastase inhibitor from equine leukocyte cytosol belongs to the  
 RT serpin superfamily. Further characterization and amino acid sequence  
 RT of the reactive center.";  
 RL J. Biol. Chem. 263:7364-7369(1988).  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (1.95 ANGSTROMS).  
 RX MEDLINE=92389327; PubMed=1518052;  
 RA Baumann U., Bode W., Huber R., Travis J., Potempa J.;  
 RT "Crystal structure of cleaved equine leucocyte elastase inhibitor  
 RT determined at 1.95-A resolution.";  
 RL J. Mol. Biol. 226:1207-1218(1992).  
 CC -1- FUNCTION: THIS INHIBITOR IS THOUGHT TO BE INVOLVED IN THE CONTROL  
 CC OF INTRACELLULAR PROTEIN TURNOVER. BINDS TO THYMOSIN BETA-4.  
 CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M91161; A097513.1; -.  
 DR PIR: A28060; A28060.  
 DR PIR: A37276; A37276.  
 DR PIR: A42421; A42421.  
 DR PIR: S34062; S34062.  
 DR PDB: 1HLE; 3I-JAN-94.  
 DR InterPro: IPR000215; Serpin.  
 DR Pfam: PF00079; serpin.1.  
 DR SMART: SM00093; SERPIN.1.  
 DR PROSITE: PS00284; SERPIN.1.  
 KM Serpin: serine protease inhibitor; 3d-structure.  
 FT MOD\_RES 1 1 BLOCKED (WITH AN UNKNOWN CHEMICAL GROUP).  
 FT ACT\_SITE 344 345  
 FT HELIX 2 22 REACTIVE BOND.  
 FT STRAND 28 30  
 FT HELIX 32 44  
 FT TURN 45 45  
 FT HELIX 48 58  
 FT TURN 59 59  
 FT HELIX 60 62  
 FT TURN 64 65  
 FT HELIX 66 77  
 FT TURN 78 78  
 FT STRAND 85 95  
 FT TURN 96 97  
 FT HELIX 102 112  
 FT STRAND 115 119  
 FT TURN 121 123  
 FT HELIX 125 139  
 FT TURN 140 142  
 FT TURN 150 151  
 FT TURN 155 156  
 FT STRAND 157 171  
 FT HELIX 177 179  
 FT STRAND 181 186  
 FT STRAND 192 209  
 FT HELIX 210 212  
 FT TURN 213 213  
 FT STRAND 214 221  
 FT TURN 222 223  
 FT STRAND 226 233  
 FT HELIX 244 248  
 FT TURN 249 249  
 FT HELIX 252 259  
 FT HELIX 261 263

FT STRAND 265 274  
 FT STRAND 276 283  
 FT HELIX 285 291  
 FT TURN 292 292  
 FT HELIX 295 297  
 FT TURN 299 301  
 FT HELIX 305 308  
 FT STRAND 313 326  
 FT STRAND 330 343  
 FT STRAND 350 353  
 FT STRAND 358 364  
 FT TURN 365 368  
 FT STRAND 369 376  
 SO SEQUENCE 379 AA; 43006 MW; 72DED1999371427A CRC64;  
 Query Match 37.7%; Score 739.5; DB 1; Length 379;  
 Best Local Similarity 38.0%; Pred. No. 1,4e-44;  
 Matches 148; Conservative 85; Mismatches 137; Indels 19; Gaps 4;  
 QY 1 MASLAANMEFCNLFREDDNNGNVPFSSLSLFAALVALVRLGAQDDSLQIDKLLHY 60  
 DB 1 MEQLSTANTHFAVDLFRALNESDPTGNIFISPLISSALAMTFLGRTAQAQVSRALYF 60  
 QY 61 NTASGYGNSSNSQSLQKRVFSDINASHKDYDSIVNGLFAEKVYGFHKDYIECAEK 120  
 DB 61 DTV-----EDHSRQSLNADINKGAPYILKLNRLYGEKTYNFLADFLASTQK 110  
 QY 121 LYDAKVRVDETHLEDFTRRNINKWVENETHGKIKVNIIGEGGISSAVMYLVNAVYFKGK 180  
 DB 111 MYGAELASVDFOQAPDARKKEINEMWCGTEGKIPPELLVGMVDNMTKLVLAIVAFKGN 170  
 QY 181 WQSAFTKSETINCHFSKCSGKAVAMHGERKFNLSVIEDPSMKITELRYNG-GINMY 239  
 DB 171 WQEKMEKENTRDPEFRLNKDKDTVKRMVTKKFPYNYIEDLKCVLELPYQKEISMTI 230  
 QY 240 LIPEN-----DLSEIENKLTFFONLMEMTPRRMTSKYVEFPQRIEKNYEMKQYLR 293  
 DB 231 LIPDDLEDESGTEKLEKQVLEKLEKMTPEKLVYLAENVHLPRKLEESYDLTJSLAR 290  
 QY 294 LGKIDFDESKADLSGIASGRLYISRMHKSITVEETSEYTAATGSIVYKQLPST 353  
 DB 291 LGVODLFNRGKADLSGMSGARDLFVSKIIHKSEVDINEGTEAATACTIMLAMLPPE 350  
 QY 354 LFRADHPFLFVIRKDDI--TLFSGKVSQP 380  
 DB 351 NFNADHPFLFIRNPSANILFLGRSSP 379  
 RESULT 7  
 ID ILEU\_HUMAN STANDARD; PRT: 379 AA.  
 AC P30740;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Leukocyte elastase inhibitor (LEI) (Monocyte/neutrophil elastase  
 DE inhibitor) (M/NEI) (EI).  
 GN SERPINB1 OR ELANH2 OR PI2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92302296; PubMed=1376927;  
 RA Remold-O'Donnell E., Chin J., Alberts M.;  
 RT "Sequence and molecular characterization of human monocyte/neutrophil  
 RT elastase inhibitor.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:5635-5639(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98296265; PubMed=9630619;  
 RA Zeng W., Silverman G.A., Remold-O'Donnell E.;

```

RT RT "Structure and sequence of human M/NEI (monocyte/neutrophil elastase
RL RL inhibitor), an ov-serpin family gene.";
RN Gene 213:179-187(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Cervix;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/Genbank/DBJ databases.
RP
RP SEQUENCE OF 178-185; 204-210 AND 364-371.
RC TISSUE=Keratinocytes;
RX MEDLINE=93162043; PubMed=1286667;
RA Rasmussen H.H., Van Damme J., Puype M., Gesser B., Celis J.E.,
RN Vanderkhorst J.;
RT "Microsequences of 145 proteins recorded in the two-dimensional gel
RL protein database of normal human epidermal keratinocytes.";
CC Electrophoresis 13:960-969(1992).
CC
CC -1- FUNCTION: REGULATES THE ACTIVITY OF THE NEUTROPHIL PROTEASES
CC ELASTASE, CATHEPSIN G AND PROTEINASE-3.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL collaboration -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/isb-ch).
CC or send an email to license@isb-slb.ch).
CC -----
DR EMBL: M93056; -1- NOT ANNOTATED_CDS.
DR EMBL: AF053630; AAC31394.1; -1-
DR EMBL: BC009015; AAH09015.1; -1-
DR PIR: S27383; S27383.
DR HSSP: P05619; 1HLE.
DR Aarhus/Ghent-2DPAGE: 4314; IEF.
DR GENE: HGNC:3311; SERPINB1.
DR MIM: 130135; -1-
DR InterPro: IPR000215; Serpin.
DR Pfam: PF00079; serpin; 1.
DR SMART: SM00093; SERPIN; 1.
DR PROSITE: PS00284; SERPIN; 1.
DR Serpin: Serine protease inhibitor.
FT ACCT_SITE 344 345 REACTIVE BOND (BY SIMILARITY).
SQ SEQUENCE 379 AA: 42741 MW: 54AE08DFBCD8CD3 CRC64;

Query Match 36.9%; Score 722.5; DB 1; Length 379;
Best Local Similarity 38.6%; Pred. No. 2,le=43;
Matches 150; Conservative 74; Mismatches 146; Indels 19; Gaps 4;

QY 1 MASLAANAWEFCNLFREMDNOCNGNVFFSSLSFALALVRLGADDSLSQIDKLLAV 60
DB 1 MEQLSANTRALDLPALFLALSENPNAGNIFISPFSSAMAMVFIQTRNAAQLSKTEHF 60
QY 61 NTAGCYGNSSQSGLOSLKRVSDINASKVDLSLVNGLFAEKYVGFHKDIIECEK 120
DB 61 NTAV-----EYHSRFQSLMADINIKKRCASYTLKANLVEKTYNLPFLVSTOK 110
QY 121 LYDAKVERVDPTNHLDETRNINKWVENETGKIKNVGEGCIGSSAAVMLVNAVFPKX 180
DB 111 TYGADLASVDFQHAASEDAKRTITNOMWKGOTGKIPELIASSAMVDMKRLVNAVTRGN 170
QY 181 WQSAFTKSETINCHFKSPKSGCAVAMAHQERKTNLSVIEDPSMKILELRVNG-GINMY 239
DB 171 WKDFEMKEATNAPFLNKKRDKTVKMMYQKKFAYGYIEDLKCRVLELTPYQGBELSVI 230
QY 240 LLPEN-----DLSEIKNLKTLFQVLMEXTNRRMTSKYVEVFFPOFKTEKVEKQYRA 293
DB 231 LLPPDIEDESTGLKIEQOLTELEKTLHEWTKPENLDLELVANSLDRPFKEESYTLNSDLAR 290
QY 294 IGLADIDFESKADISLGSAGRLKISRMHMSYLEVMEEGTEAANAAGSNVVEKQLPOST 353
DB 291 LGVDDLFSSKADISLGSAGARDITISKVHSEFVEVNEEGTEAANAAGTATPCMLPDEE 350

```

DB	351	NTFADHPLEFIRHNSSGSLIFLGFSSP	379
OY	354	LEFRADHPLEFVIRKDD--ILIFSGVSCP	380
DB	351	NTFADHPLEFIRHNSSGSLIFLGFSSP	379
RESULT	8		
ID	SB13_HUMAN	STANDARD:	PRT; 391 AA.
AC	SB13_HUMAN	Q9UBW1; Q9UKG0; Q9HCX1;	
DT	16-OCT-2001	(Rel. 40, Created)	
DT	15-OCT-2001	(Rel. 40, Last sequence update)	
DT	15-JUN-2002	(Rel. 41, Last annotation update)	
DE	Hurpin (HACar UV-repressible serpin) (Protease inhibitor 13)		
DE	(Headpin) (Serpin B13).		
GN	SERPINA13 OR PIL3.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]	SEQUENCE FROM N.A. (ISOFORM 1).	
RC	TISSUE=Skin;		
RX	MEDLINE=99443898; PubMed=105127713;		
RA	Abts H.F., Welss T., Mirmohammadsadeh A., Koehrer K., Michel G.,		
RA	Ruzicka T.;		
FT	"Cloning and characterization of hurpin (Protease inhibitor 13): a new		
FT	skin specific, UV-repressible serine proteinase inhibitor of the		
FT	ovalbumin serpin family".		
RL	J. Mol. Biol. 293:29-39(1999).		
RN	[2]	SEQUENCE FROM N.A. (ISOFORM 1).	
RP	MEDLINE=99458661; PubMed=10527881;		
RX	Spring P., Nakashima T., Frederick M., Henderson Y., Clayman G.;		
RA	"Identification and cDNA cloning of headpin, a novel differentially		
RT	expressed serpin that maps to chromosome 16q."		
RL	Biochem. Biophys. Res. Commun. 264:299-304(1999).		
RN	[3]	SEQUENCE FROM N.A. (ISOFORM 1).	
RP	MEDLINE=20461850; PubMed=11004515;		
RX	Nakashima T., Pak S.C., Silverman G.A., Spring P.M., Frederick M.J.,		
RA	Clayman G.L.;		
RT	"Genomic cloning, mapping, structure and promoter analysis of HEADPIN,		
RT	a serpin which is down-regulated in head and neck cancer cells."		
RL	Biochim. Biophys. Acta 1492:441-446(2000).		
RN	[4]	SEQUENCE FROM N.A. (ISOFORM 2).	
RP	TISSUE=Placenta;		
RC	MEDLINE=21213570; PubMed=11313015;		
RX	Abts H.F., Welss T., Scheuring S., Scott F.L., Irving J.A., Michel G.,		
RA	Bird P.I., Ruzicka T.;		
RT	"Sequence, organization, chromosomal localization and alternative		
RT	splicing of the human serine proteinase inhibitor gene hurpin (PIL3),		
RT	which is up-regulated in psoriasis."		
RL	DNA Cell Biol. 20:123-131(2001).		
CC	-1- FUNCTION: May play a role in the proliferation or differentiation		
CC	of keratinocytes.		
CC	-1- SUBCELLULAR LOCATION: Cytoplasmic.		
CC	-1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; are		
CC	produced by alternative splicing.		
CC	-1- TISSUE SPECIFICITY: SKIN-SPECIFIC.		
CC	-1- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation in the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announcement">http://www.isb-sib.ch/announcement</a> ).		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> .		
EMBL	AI001696; CA004935.2; -		
EMBL	AI001697; CA004936.2; -		

```

DR EMBL: AJ001698; CAA04937.1; -
DR EMBL: AF169949; AAD5765.1; -
DR EMBL: AF216854; AAF27879.1; -
DR EMBL: A278717; CAC03569.1; -
DR HSSP: P05120; 1BY7.
DR Genew; HGNC:8944; SERPINB13.
DR MIM; 604445; -
DR InterPro: IPR000240; Maspin.
DR InterPro: IPR000215; Serpin.
DR Pfam: PF00079; serpin.1.
DR PRINTS; PR00676; MASPIN.
DR SMART; SM00093; SERPIN.1.
DR PROSITE; PS00284; SERPIN; 1.
DR Serpin; Serine protease inhibitor. Alternative splicing.
KM ACT_SITE 356 357 REACTIVE_BOND (BY SIMILARITY).
FT VARSPIC 206 257 MISSING (IN ISOFORM 2).
FT CONFLICT 8 8 S -> N (IN REF. 4).
FT CONFLICT 75 75 MISSING (IN REF. 1; CAA04937).
FT CONFLICT 293 293 G -> S (IN REF. 2).
FT CONFLICT 297 297 E -> Q (IN REF. 1; CAA04937).
SQ SEQUENCE 391 AA; 44276 MW; 2CA8858D4BC2B09 CRC64;

Query Match 36.5%; Score 714.5; DB 1; Length 391;
Best Local Similarity 38.0%; Pred. No. 7, 9e-43;
Matches 152; Conservative 77; Mismatches 142; Indels 29; Gaps 7;

OY 1 MASLAANAEEFCNLFREMDNONGNNGVFFSSLSLFAALLVLRGAODDSLQIDKLHV 60
DB 1 MDSLGAVSTRGLGDLFEEL-KKTNDGNIFSPVGLTALIGVLTGRGATASOLEEYFH- 58
OY 61 NTASGVGNSSNSGSLQSLKRVFSDINASHK-----DYDSIVNGLPFAEK 106
DB 59 -----SEKTKSSRIKAEKEVIENTEAVHOQFOKFLTISKLTNDYELINTNRLEGEK 112
OY 107 VYGFHKDYIECAEKLYDAKVERVDFTNHLDETNRINKWVENETHGKIKVIGEGISS 166
DB 113 TYLFLOKYDYVEKYYHASELEPVDFVNADESRRKIKSVESKTEKIKDLFPDGSISS 172
OY 167 AVAVLVAVYFKGKMOGSAFTKSEINCHFKSPKCSGKAVAMMOERKNLSVIEDPKMI 226
DB 173 TKLVLMNVYFKGMDREFFKENTKEKFMNKSSTKSVOMQOSHSFTFLFDLOAKI 232
OY 227 LEDRY-NGGIMVYLLPNDLSEIE--NKLTFQNLMEWTPRRMTSKYVEVFPQRIE 282
DB 233 LGIPYKNDLSMFLLP-NDIDGLEKIDIKISPEKLEWTPSPGMEERKYNLHLPREVE 291
OY 283 KNEEMKOYLALGLKIDFDSKADLSGIAGGRLYISRMHKSITYEVEGTEATATGS 342
DB 292 DGYDLAVLVAMGMDAFSEHKADYSGMSGSLYAOGLFHSFVAATEEGTEAATAATGI 351
OY 343 NIVEKQLPOSTLFRADHPFLFYIR--KDIILFSGKVSPP 380
DB 352 GFTVTSAPGHENYCHNHPFLFTRHNESNLSLFFGRPSSP 391

RESULT 9
OVARY_CHICK STANDARD: PRT; 388 AA.
AC P01014;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE Gene y protein (Ovalbumin-related).
GN y
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauilia; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=63014329; Pubmed=7122240;
RA Hellig R., Muraskovsky R., Kioepfer C., Mandel J.L.;

```

```

RT "The ovalbumin gene family: complete sequence and structure of the y
RT gene.";
RL Nucleic Acids Res. 10:4363-4382(1982).
CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: J00922; AAA68882.1; -
DR PIR; A01244; DYCH.
DR HSSP; P01012; IOVA.
DR InterPro: IPR000215; Serpin.
DR Pfam; PF00079; serpin.1.
DR SMART; SM00093; SERPIN.1.
DR PROSITE; PS00284; SERPIN; 1.
KW Serpin.
KM ACT_SITE 353 354 REACTIVE_BOND_HOMOLOG.
FT CARBOHYD 95 95 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 215 215 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 312 312 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 388 AA; 43772 MW; 2AF17BF150F461 CRC64;

Query Match 36.3%; Score 711; DB 1; Length 388;
Best Local Similarity 39.6%; Pred. No. 1, 4e-42;
Matches 155; Conservative 75; Mismatches 147; Indels 14; Gaps 7;

OY 1 MASLAANAEEFCNLFREMDNONGNNGVFFSSLSLFAALLVLRGAODDSLQIDKLHV 60
DB 1 MDSISVINAEEFCEDVFNEMVHHVHNENILCPISLIALMAYLVLRGAGNSESQMKVLYF 60
OY 61 NTASGVGNSSNSGSLQSLKRVFSDINASHKDYDSIVNGLPFAEKYVGFHKDYIE 116
DB 61 DSITGAGSTIDSGGSEYVHNLEKELSETTRNATYSLIEDKLYVDFTFSLPYLS 120
OY 117 CAEKLYDAKVERVDFTNHLDETNRINKWVENETHGKIKVIGEGISSAVVAVNAVY 176
DB 121 CARFYYGVGEVVEVFKAAEAEARQLINSWEKETNGOIKDLVSSSIDFTWVFINTIY 180
OY 177 FKGMOSAFTRKSEINCHFKSPKCSGKAVAMMOERKNLSVIEDPKMLERY-NGGI 235
DB 181 FKGIWKIAETEDTREMPEMTKEESKPVOMCMNNNSFNATVLAERMKLTLEPYASGDL 240
OY 236 NMVYLPE--NDLSEIENKLTFLQNLMEWTPRRMTSKYVEVFPQRIENYEMKOYLRA 293
DB 241 SMLVLPDDEVSGLEIREKTINFDKLRMTSTNMMAKSMKVIYLPKMIIEKYNLTSLMA 300
OY 294 LGLKIDFDSKADLSGIAGGRLYISRMHKSITYEVEGTEATATGS--NIVEKQLPQ 351
DB 301 LGMTDLFSRS-ANLTGISSVDNLMISDAVGEVEVEEVEGTEATGSAIGNI--KHSLE 357
OY 352 STLFRAADHPFLFYIRKD--DIILFSGKVSPP 380
DB 358 LEEFRADHPFLFYIRNPTNALIFFGRYWSPP 388

RESULT 10
PT16_BOVIN STANDARD: PRT; 378 AA.
AC O02739;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Serine proteinase inhibitor B-43.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.

```

OX NCBI\_TaxID=9913;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=97165893; PubMed=9013786;  
 RA Nakaya N., Nishihori M., Kawabata M., Saeki K.;  
 RT "Cloning of a serine proteinase inhibitor from bovine brain:  
 expression in the brain and characterization of its target  
 proteinases.";  
 RL Brain Res. Mol. Brain Res. 42:293-300(1996).  
 CC -1- FUNCTION: MAY BE INVOLVED IN THE REGULATION OF SERINE PROTEINASES  
 CC PRESENT IN THE BRAIN OR EXTRASARIAL FROM THE BLOOD.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- TISSUE SPECIFICITY: BRAIN.  
 CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: D55670; BAA19875.1; -  
 DR HSSP: P05120; 1BY7.  
 DR InterPro: IPR000215; Serpin.  
 DR Pfam: PF00079; serpin.1.  
 DR SMART: SM00093; SERPIN: 1.  
 DR PROSITE: PS00284; SERPIN: 1.  
 DR Serpin; Serine protease inhibitor.  
 FT ACT\_SITE 343 344  
 FT SEQUENCE 378 AA; 42560 MW; 664F499CCCE263A CRC64;  
 SQ  
 Query Match 35.8%; Score 702; DB 1; Length 378;  
 Best Local Similarity 39.4%; Pred. No. 5.6e-42;  
 Matches 152; Conservative 83; Mismatches 137; Indels 14; Gaps 7;  
 OY 1 MASLAANAEEFCNLFREMDNONGN-NYFFSLSLFALALVRLGADDSLSQIDKLH 59  
 DB 1 MDLSANGTFLTLTKL-GEQNSKNVFISPLSISALAVLLGAKNTAAQMGQTLS 58  
 OY 60 VNTASGYGNSNSQSGLSQKRVFSDINASHKVDLSVNGFAEKVYGFHKDYECAR 119  
 DB 59 LKSSGSG--EDVHQGFQNL-----SEVNRDQYLLRTANRLFGKTYDFLSSFDSC 112  
 OY 120 KLYDAKVERVDFTNHEDRNRINKVWENETHGKIKVNGEGISSAAMVAVNAVYFKG 179  
 DB 113 KRYQAMEELDYVSATQSRKHNTWAEKTEKIDLLPANSVNPRLVLVNAVYFKG 172  
 OY 180 KWQSAFTKSETINCHKSPKSGKAVAMNHQERKFNLSYIEDPSMKILELRVNG-GINNY 238  
 DB 173 NMDTQENKHEEPRFVRSKVKNEKPYQMFKSTCKITYIGELISQIDLLPVVGQELNV 232  
 OY 233 VLLP--ENDLSEIKLTFQNLMEWTNPRMTSKYVEVFPOFKIEKNTKMOYLRAQL 296  
 DB 233 ILPESBDLNTVEALYEFKFIAMTKPDVMDDEEVEVLPRLTEESYDMEEFLOELGM 292  
 OY 297 KLIPEESKADLSGIASGRLYISRMMHKSIEVTEGTEATAGTSNIYEKOLPOSTLFR 356  
 DB 293 TPAFEETRADFSGMSSGRGLHLSKVHKSFEVTEGTEAATAAGVAMRCLMVPVPRN 352  
 OY 357 ADHPFLFYIR--KDIILFSGKVSQP 380  
 DB 353 ANHPFLFYIRHNSKGTALIFCGRFCSF 378  
 RESULT 11  
 ID ILEU\_PIG STANDARD; PRT; 378 AA.  
 AC P80229;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE leukocyte elastase inhibitor (LEI) (Leukocyte neutral proteinase  
 DE inhibitor) (LNP1).  
 GN SERPINB1 OR ELANH2.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.  
 OX NCBI\_TaxID=9823;  
 RN (1)  
 RP SEQUENCE.  
 RC TISSUE=Leukocyte;  
 RX MEDLINE=94039085; PubMed=7901009;  
 RA Teschauer W.F., Mentele R., Sommerhoff C.P.;  
 RT "Primary structure of a porcine leukocyte serpin.";  
 RL Eur. J. Biochem. 217:519-526(1993).  
 CC -1- FUNCTION: REGULATES THE ACTIVITY OF THE NEUTROPHIL PROTEASE  
 CC ELASTASE, CATHESPIN G AND PROTEINASE-3 (By similarity).  
 CC -1- SUBUNIT: HOMODIMER; DISULFIDE-LINKED.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.  
 DR HSSP: P05619; 1HLE.  
 DR InterPro: IPR000215; Serpin.  
 DR Pfam: PF00079; serpin.1.  
 DR SMART: SM00093; SERPIN: 1.  
 DR PROSITE: PS00284; SERPIN: 1.  
 DR Serpin; Serine protease inhibitor.  
 FT MOD\_RES 1 1  
 FT DISULFID 80 80  
 FT ACT\_SITE 343 344  
 FT SEQUENCE 378 AA; 42512 MW; 3CA7C65F11E508CA CRC64;  
 SQ  
 Query Match 35.5%; Score 696; DB 1; Length 378;  
 Best Local Similarity 37.5%; Pred. No. 1.5e-41;  
 Matches 146; Conservative 82; Mismatches 141; Indels 20; Gaps 5;  
 OY 1 MASLAANAEEFCNLFREMDNONGN-NYFFSLSLFALALVRLGADDSLSQIDKLH 60  
 DB 1 MDLSAANRFDALDFRALNESNPAGNIFISPSISALAMILLGTGRTGAQMSKALHF 60  
 OY 61 NTASGYGNSNSQSGLSQKRVFSDINASHKVDLSVNGFAEKVYGFHKDYECAR 120  
 DB 61 DTV-----KDIHSFGSLADINKGASYLKLANRLEFKTYFLPEFLASQK 110  
 OY 121 LYDAKVERVDFTNHEDRNRINKVWENETHGKIKVNGEGISSAAMVAVNAVYFKG 180  
 DB 111 TYGAELASVDPLRASBEAKKAINEMWBEQTEKIDELLASGVDSATKLVNAVYFKGS 170  
 OY 181 WQSAFTKSETINCHKSPKSGKAVAMNHQERKFNLSYIEDPSMKILELRVNG-GINNY 239  
 DB 171 WQEKMTATKDAAPRLNKKDSKTYMMYQKRRFPFGYIKELKCRVLELPYQKDLNVI 230  
 OY 240 LLPEN-----DLSEIKLTFQNLMEWTNPRMTSKYVEVFPOFKIEKNTKMOYLRA 293  
 DB 231 LLPDSIEDSTGLRKITEQHTLEKLNEMTKPRLNLELVNAVLPFRLESDIADLAPLR 290  
 OY 294 IGLKIDFESKADLSGIASGRLYISRMMHKSIEVTEGTEATAGTSNIYEKOLPOST 353  
 DB 291 LGVDDLEF--SRADLGMSEARDLFIISKYVHKAFVEVNEGTEAATAAGVAMRCLMPEE 349  
 OY 354 LFRADHPFLFYIRKD--DIILFSGKVSQP 380  
 DB 350 DFIADHPFLFYIRHNPSSNILELGRILSP 378  
 RESULT 12  
 ID PAI2\_HUMAN STANDARD; PRT; 415 AA.  
 AC P05120; Q96E96;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE plasminogen activator inhibitor-2 precursor (PAI-2) (Placental  
 DE plasminogen activator inhibitor) (Monocytic Arg-serpin) (Urokinase



DE inhibitor).  
GN SERPINB2 OR PAI2 OR PLANH2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC TISSUE-Placenta;  
RX MEDLINE=87137674; PubMed=3029122;  
RA Ye R.D., Mun T.-Z., Sadler J.E.;  
RT "CDN cloning and expression in Escherichia coli of a plasminogen  
RT activator inhibitor from human placenta.";  
RL J. Biol. Chem. 263:3718-3725(1987).  
RN (2)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88142852; PubMed=3325828;  
RA Schleuning W.-D., Medcalf R.L., Hession C., Rothenbuhler R.,  
RA Shaw A., Krulthof E.K.O.;  
RT "Plasminogen activator inhibitor 2: regulation of gene transcription  
RT during phorbol ester-mediated differentiation of U-937 human  
RT histiocytic lymphoma cells.";  
RL Mol. Cell. Biol. 7:4564-4567(1987).  
RN (3)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89174589; PubMed=2494165;  
RA Ye R.D., Ahern S.M., Le Beau M.M., Lebo R.V., Sadler J.E.;  
RT "Structure of the gene for human plasminogen activator inhibitor-2.  
RT The nearest mammalian homologue of chicken ovalbumin.";  
RL J. Biol. Chem. 264:5495-5502(1989).  
RN (4)  
RP SEQUENCE FROM N.A.  
RC TISSUE-Monocytes;  
RX MEDLINE=87252928; PubMed=3496414;  
RA Webb A.C., Collins K.L., Snyder S.F., Alexander S.J.,  
RA Rosenwasser L.J., Eddy R.L., Shows T.B., Auron P.E.;  
RT "Human monocyte Arg-Serpin cDNA. Sequence, chromosomal assignment,  
RT and homology to plasminogen activator-inhibitor.";  
RL J. Exp. Med. 166:77-94(1987).  
RN (5)  
RP SEQUENCE FROM N.A.  
RC TISSUE-Monocytes;  
RX MEDLINE=88125032; PubMed=3257578;  
RA Antalis T.M., Clark M.A., Barnes T., Lehtbach P.R., Devine P.L.,  
RA Schevov G., Goss N.H., Stephens R.W., Tolstoshev P.;  
RT "Cloning and expression of a cDNA coding for a human monocyte-derived  
RT plasminogen activator inhibitor.";  
RL Proc. Natl. Acad. Sci. U.S.A. 85:985-989(1988).  
RN (6)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90152678; PubMed=2303256;  
RA Simla J.A., Alexander S.J., Horton K.W., Auron P.E., Byers M.G.,  
RA Shows T.B., Jr., Webb A.C.;  
RT "Chromosomal organization and localization of the human urokinase  
RT inhibitor gene: perfect structural conservation with ovalbumin.";  
RL Genomics 6:159-167(1990).  
RN (7)  
RP SEQUENCE FROM N.A.  
RC TISSUE-Brain;  
RX Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.  
RN (8)  
RP SEQUENCE OF 12-17, 103-108 AND 314-321.  
RX MEDLINE=93162043; PubMed=1286667;  
RA Resmussen H.H., van Damme J., Puyse M., Gesser B., Celis J.E.,  
RA Vandeckerckhove J.;  
RT "Microsequences of 145 proteins recorded in the two-dimensional gel  
RT protein database of normal human epidermal keratinocytes.";  
RL Electrophoresis 13:960-969(1992).  
RN (9)  
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
RX MEDLINE=9918112; PubMed=10368272;  
RA Harrop S.J., Jankova L., Coles M., Jardine D., Whittaker J.S.,

RA Could A.R., Meister A., King G.C., Mabbutt B.C., Curmi P.M.G.;  
RT "The crystal structure of plasminogen activator inhibitor 2 at 2.0-A  
RT resolution: implications for serpin function.";  
RL Structure 7:43-54(1999).  
CC CC THE MONOCYTE DERIVED PAI-2 IS DISTINCT FROM THE ENDOTHELIAL  
CC CELL-DERIVED PAI-1.  
CC -1- SUBCELLULAR LOCATION: CYTOSOL; PLASMA MEMBRANE; EXTRACELLULAR.  
CC -1- PPM: THE SIGNAL SEQUENCE IS NOT CLEAVED.  
CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: J02685; AAA36413.1; -;  
DR EMBL: M18082; AAA60006.1; -;  
DR EMBL: Y00630; CAA68666.1; -;  
DR EMBL: M24657; AAA60348.1; -;  
DR EMBL: M24651; AAA60348.1; JOINED.  
DR EMBL: M24652; AAA60348.1; JOINED.  
DR EMBL: M24653; AAA60348.1; JOINED.  
DR EMBL: M24654; AAA60348.1; JOINED.  
DR EMBL: M24655; AAA60348.1; JOINED.  
DR EMBL: M24656; AAA60348.1; JOINED.  
DR EMBL: J03603; AAA60004.1; -;  
DR EMBL: M31551; AAA36797.1; -;  
DR EMBL: M31547; AAA36797.1; JOINED.  
DR EMBL: M31548; AAA36797.1; JOINED.  
DR EMBL: M31549; AAA36797.1; JOINED.  
DR EMBL: M31550; AAA36797.1; JOINED.  
DR EMBL: BC012609; AAH12609.1; -;  
DR EMBL: A21238; CAA01535.1; -;  
DR EMBL: A21254; CAA01539.1; -;  
DR PIR: A26553; A26553.  
DR PIR: A29362; A29362.  
DR PIR: A32853; A32853.  
DR PIR: A31366; A31366.  
DR PDB: 1BY7; 24-OCT-99.  
DR Aarhus/Chent-2DPAGE; 6314; IEF.  
DR Genew; HGNC:8584; SERPINB2.  
DR MIM; 173390; -;  
DR InterPro; IPR000215; Serpin.  
DR Pfam; PF00079; serpin.1.  
DR SMART; SM00093; SERPIN.1.  
DR PROSITE; PS00284; SERPIN.1.  
KW Serpin; Serine protease inhibitor; Plasma; Plasminogen activation;  
KW Glycoprotein; Signal; 3D-structure; Polymorphism.  
FT SIGNAL 1 ?  
FT CARBOHYD 75 75 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 115 115 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 339 339 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT ACT\_SITE 380 381 REACTIVE BOND.  
FT VARIANT 120 120 N -> D (IN DBSNP: 6098).  
FT VARIANT 404 404 /FTID=VAR.011743.  
FT VARIANT 413 413 N -> K (IN DBSNP: 6103).  
FT VARIANT 413 413 S -> C (IN DBSNP: 6104).  
FT CONFLICT 170 170 /FTID=VAR.011745.  
FT SEQUENCE 415 AA; 46596 MW; 10DFAB5A4B1246FF CRC64;  
SQ  
Query Match 33.9%; Score 664.5; DB 1; Length 415;  
Best Local Similarity 34.7%; Pred. No 2.6e-39;  
Matches 144; Conservative 83; Mismatches 153; Indels 35; Gaps 5;  
Qy 1 MASIAANAEFCENFREMNDGNGCNVFFSSLSLFAALVRLGAQDSDLQIDRLHY 60  
Db 1 MEDLCVANTLFAALNLFRLAKASPTQNLFLSPMSISSTAMVYVGRSGSTEDQMAVLOF 60

QY 61 N-----TASGY-----GN-----SSNSGSLQOLKRVFSDINASHD 93  
 DB 61 NEGANAAVTPMTPENTSCGFMQIQKGSVPAILQAQADKHHSSFRSLSSAINSTGN 120  
 QY 94 YDLSTVNGFAEKVYGFHHDYIECAKLYDAKVERDFNHLNEDTRRNINKVNEETHGK 153  
 DB 121 YLLESTVNGFAEKVYGFHHDYIECAKLYDAKVERDFNHLNEDTRRNINKVNEETHGK 180  
 QY 154 INKIVGEGGSSAVVAVLVNAVYFKGMQSAFTKSEFTINCHFPKSPKCAVAMHMQEKK 213  
 DB 181 INNLPEGSVDGDTFRAVLNAVYFKGMQSAFTKSEFTINCHFPKSPKCAVAMHMQEKK 240  
 QY 214 FNLSTVEDPSMKILELRNGINMYVLPE-----NDLSEIENKLTFFONLMEWTPRRM 267  
 DB 241 LNVGYEDLKAQIILELPYAGDVSMFLLPDEIADVSTGLLESEITTYDKLNKMTSKDM 300  
 QY 266 TSKYVVEFPFKIEKNYEMKOYLALGLKIDFDESKADLSGASGRLYISMAMKSTI 327  
 DB 301 AEDEVEVYIPORLEHVELRSLRSMGMDAFNKRANFGSGSERNDLFLSEVFHQAVY 360  
 QY 328 EYTEEGTEATNGSNIVEKQLPOSTLFRADHPFLFVI--RKDDILFSGKVS 380  
 DB 361 DVNEEGTEAAGTGGVMTGRTGHGPGQVADHPFLFLMHKTTNCLTFGRSSP 415

## RESULT 13

ID SPB8\_HUMAN STANDARD: PRT: 374 AA.  
 AC P50452:  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DE 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Cytoplasmic antiprotease 2 (CAP2) (CAP-2) (Protease inhibitor 8) (Serpin B8).  
 GN SERPINB8 OR P18.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=96102039; PubMed=8530382;  
 RA Sprecher S.A., Morgenstern K.A., Mathews S., Dahlen J.R.,  
 RA Schrader S.K., Foster D.C., Kistel W.;  
 RT "Molecular cloning, expression, and partial characterization of two novel members of the ovalbumin family of serine proteinase inhibitors";  
 RT J. Biol. Chem. 270:29854-29861(1995).  
 RL J. Biol. Chem. 270:29854-29861(1995).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL: L40377; AAC1939.1; -  
 DR HSSP: P05120; 1B77.  
 DR Genew: HGNC:8952; SERPINB8.  
 DR MIM: 601697; -  
 DR InterPro: IPR000215; Serpin.  
 DR Pfam: PF00079; serpin.1.  
 DR SMART: SM00093; SERPIN.1.  
 DR PROSITE: PS00284; SERPIN.1.  
 KW Serpin; Serine protease inhibitor.  
 FT ACT SITE 339 340 REACTIVE BOND (BY SIMILARITY).  
 SQ SEQUENCE 374 AA. 42786 MW: E855A03519AC60E CRC64;

Query Match 33.8%; Score 662; DB 1; Length 374;  
 Best Local Similarity 37.4%; Pred. No. 3.4e-39;  
 Matches 145; Conservative 79; Mismatches 142; Indels 22; Gaps 8;

QY 1 MASLAANAEPENLFREDDNNGNVEFSSLSFALALVRLGAQDLSQIDKLHV 60  
 DB 1 MDLCEANCTFAISLEKILGEEEDNSRNVFSPMSISALAMVFGKSGSTAOMSQL 60  
 QY 61 NTASGNGSNSOSGSLQOLKRVFSDINASHDIDYIYNGFAEKVYGFHHDYIECAK 120  
 DB 61 Y-----KGGDHLRGQSLL-----SEVNTGTOYLRTNRRFLGETCTCFDLDFKEXCK 110  
 QY 121 LYDAKVERDFNHLNEDTRRNINKVNEETHGKIKKIVGEGGSSAVVAVLVNAVYFKG 180  
 DB 111 FYQAELEELSPADEDEECHKHINDVAEETEKISEVLDAGTVDP LTKVLVNAVYFKG 170  
 QY 181 WQSAFTKSEFTINCHFPKSPKCAVAMHMQEKNLSVIEDSMKILEARY-NGGINMYV 239  
 DB 171 WNEQDFRKYTRKMLEFTNE-EKRYVQMPKREKFKGYADVHTVLELPYEEELSVYI 229  
 QY 240 LLPEN--DLSEIENKLTFFONLMEWTPRRMTSKYVEVFPFKIEKNYEMKOYLALGLK 297  
 DB 230 LLPDQNTDLAVVEKALTYEKFKAWTNSEKLTSGVQLPRKLESTYDLPEFLRLGWI 289  
 QY 298 DIFDESKADLSGASGRLYISMAMKSTIYETEGTEATNG---SNIVEKQLPOSTL 354  
 DB 290 DAFDEAKADFGSMSTENKYNPLSKVAHCKEVEVNEEGTEAATAAVVRNRCRMEPR--- 346  
 QY 355 FRADHPFLFVIR--KDDILFSGKVS 380  
 DB 347 FCADHPFLFVIRHHTNCLTFGRSSP 374

## RESULT 14

ID SB10\_HUMAN STANDARD: PRT: 397 AA.  
 AC P48595;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DE 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Bomanin (Protease inhibitor 10) (Serpin B10).  
 GN SERPINB10 OR P110.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Bone marrow;  
 RX MEDLINE=96070759; PubMed=7592909;  
 RA Riewald M., Schlegel R.R.;  
 RT "Molecular cloning of bomanin (protease inhibitor 10), a novel human serpin that is expressed specifically in the bone marrow";  
 RT J. Biol. Chem. 270:26754-26757(1995).  
 RL J. Biol. Chem. 270:26754-26757(1995).  
 CC -1- FUNCTION: MAY PLAY A ROLE IN THE REGULATION OF PROTEASE ACTIVITIES DURING HEMATOPOIESIS.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN THE BONE MARROW.  
 CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL: U35459; AAC50282.1; -  
 DR HSSP: P05619; 1HLE.  
 DR Genew: HGNC:8942; SERPINB10.  
 DR MIM: 602058; -  
 DR InterPro: IPR000215; Serpin.

DR Pfam: PF00079; serpin; 1.  
 DR SMART: SM00093; SERPIN; 1.  
 DR PROSITE: PS00284; SERPIN; 1.  
 KW Serpin; Serine protease inhibitor.  
 FT ACT\_SITE 362 363 REACTIVE\_BOND (BY SIMILARITY).  
 SQ SEQUENCE 397 AA: 45402 MW: 8CE0124686715ADF CRC64;

Query Match 33.6%; Score 658.5; DB 1; Length 397;  
 Best Local Similarity 34.9%; Pred. No. 6,5e-39;  
 Matches 139; Conservative 89; Mismatches 151; Indels 19; Gaps 5;

QY 1 MASLAANAEFCENLEFREMDDNGNCNVFSSLSLEPAALALVGLAODSLQIDKLHV 60  
 D 1 MSLATSIQFALESLKLAESAGKNIFSSMSISLTLYLVGAKGTAAQAOVLQF 60  
 QY 61 NIASGCG-----NSNSQSGLOLKRVPFSDIMASHKTYDLSIVGLAEKY 107  
 D 61 NRDOGVKCDPESEKRRKMEFNLSNSEE-IHSDFOTLISELTKENDDYLLKTANAIVGEKT 119  
 QY 108 YGFHKDYICAEKLYDAKVERVDFTNHLIEDTRINKWVENETHGKIKNYIGGGISSA 167  
 D 120 YAFHNKLYEDMKTYFAEPQPVVFASDQIRKDISWVEROTEKIQNLPPDSVDSTT 179  
 QY 168 VAVLVNAVYFKGKMQSAFTKSETINCHFKSPKSGRAVAMMHQERKFNLSVIDEPPSKIL 227  
 D 180 RMLVNAVLYFKGIWEHQFLVQNTTEKPRINETTSKPYQMMFKKHLFIHIEKPRAYGL 239  
 QY 228 ELRYNG-GINMYVLDE--NDLSEINKLTPONLMETNPRRNTSKYVEVEFPQFIENK 284  
 D 240 QLYKSRDLSLLLPEDINGLEOLEKATYEKLNEMTSADNMELEYEVOLHLPKFLIEDS 299  
 QY 285 YEMKQYLRALGLKIDESKADLSGASGRLYISRMHKSLEYEVEEGEATTAATGSNI 344  
 D 300 YDLSKTLSSKMSDASOSKADSGMSANLFLSNVFAFVEINDEGTEAAGSGSEI 359  
 QY 345 VEKOLPQSTLFRADHPLEFVIR--KDDIILFSGKVSCEP 380  
 D 360 DIRIRVPISIEFNANHPFLFFIRHNKNTILFYGRLCSP 397

RESULT 15  
 OVAL\_CHICK  
 ID OVAL\_CHICK STANDARD; PRT; 385 AA.  
 AC P01012;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUN-2002 (Rel. 01, Last annotation update)  
 DE Ovalbumin (plakalbumin) (Allergen Gal d 2) (Gal d II).  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OC NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=82069038; PubMed=6272839;  
 RA Moo S.L.C., Beattie W.G., Catterall J.F., Dugaiczky A., Staden R.,  
 RA Brownlee G.G., O'Malley B.W.;  
 RT "Complete nucleotide sequence of the chicken chromosomal ovalbumin  
 RT gene and its biological significance.";  
 RL Biochemistry 20:6437-6446(1981).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=78199842; PubMed=661981;  
 RA McCreynolds L., O'Malley B.W., Nisbet A.D., Fothergill J.E., Givol D.,  
 RA Fields S., Robertson M., Brownlee G.G.;  
 RT "Sequence of chicken ovalbumin mRNA.";  
 RL Nature 273:723-728(1978).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=79010682; PubMed=692731;  
 RA Catterall J.F., O'Malley B.W., Robertson M.A., Staden R.,  
 RA Tanaka Y., Brownlee G.G.;

RT "Nucleotide sequence homology at 12 intron-exon junctions in the  
 RT chick ovalbumin gene.";  
 RL Nature 275:510-513(1978).  
 RN [4]  
 RP SEQUENCE OF 1-35.  
 RX MEDLINE=78116057; PubMed=272676;  
 RA Palmiter R.D., Gagnon J., Maish K.A.;  
 RT "Ovalbumin: a secreted protein without a transient hydrophobic leader  
 RT sequence.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 75:94-98(1978).  
 RN [5]  
 RP SEQUENCE OF 1-16.  
 RX MEDLINE=79186958; PubMed=751625;  
 RA Thompson E.O.P., Fisher W.K.;  
 RT "A correction and extension of the acetylated amino terminal sequence  
 RT of ovalbumin.";  
 RL Aust. J. Biol. Sci. 31:443-446(1978).  
 RN [6]  
 RP SEQUENCE OF 5-16; 29-35; 60-78; 115-123; 366-373 AND 379-385.  
 RX MEDLINE=79186957; PubMed=751624;  
 RA Thompson E.O.P., Fisher W.K.;  
 RT "Amino acid sequences containing half-cysteine residues in ovalbumin.";  
 RL Aust. J. Biol. Sci. 31:433-442(1978).  
 RN [7]  
 RP X-RAY CRYSTALLOGRAPHY (1.95 ANGSTROMS).  
 RX MEDLINE=90370102; PubMed=2395463;  
 RA Stein P.E., Leslie A.G.W., Finch J.T., Turnell W.G., McLaughlin P.J.,  
 RA Carrell R.W.;  
 RT "Crystal structure of ovalbumin as a model for the reactive centre of  
 RT serpins.";  
 RL Nature 347:99-102(1990).  
 RN [8]  
 RP X-RAY CRYSTALLOGRAPHY (1.95 ANGSTROMS).  
 RX MEDLINE=92046044; PubMed=1942038;  
 RA Stein P.E., Leslie A.G.W., Finch J.T., Carrell R.W.;  
 RT "Crystal structure of uncleaved ovalbumin at 1.95-A resolution.";  
 RL J. Mol. Biol. 221:941-959(1991).  
 RN [9]  
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).  
 RX MEDLINE=90278960; PubMed=2352279;  
 RA Wright H.T., Qian H.X., Huber R.;  
 RT "Crystal structure of plakalbumin, a proteolytically nicked form of  
 RT ovalbumin. Its relationship to the structure of cleaved alpha-1-  
 RT proteinase inhibitor.";  
 RL J. Mol. Biol. 213:513-528(1990).  
 RN [10]  
 RP REVIEW.  
 RX MEDLINE=21312433; PubMed=11419711;  
 RA Huntington J.A., Stein P.E.;  
 RT "Structure and properties of ovalbumin.";  
 RL J. Chromatogr. B 756:189-198(2001).  
 CC -1- FUNCTION: Not known.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: Major protein of egg white.  
 CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.  
 CC -1- DATABASE: NAME=worthington enzyme manual;  
 CC WWW="http://www.worthington-biochem.com/manual/OA.html".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: J00895; AAB59956.1; -;  
 DR EMBL: V00438; CAA23716.1; -;  
 DR EMBL: V00383; CAA23682.1; -;  
 DR EMBL: M34352; AAA48998.1; -;  
 DR EMBL: M34346; AAA48998.1; JOINED.  
 DR EMBL: M34347; AAA48998.1; JOINED.  
 DR EMBL: M34348; AAA48998.1; JOINED.



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: July 11, 2003, 11:44:31 ; Search time 165.904 Seconds  
(without alignments)  
471.946 Million cell updates/sec

Title: US-10-091-442-34  
Perfect score: 1959  
Sequence: 1 MASLAANAFCFNLFRMD.....FLFVKRDIILFSGKVSQP 380

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues  
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_protist:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1505	76.8	380	11 Q9D695	Q9D695 mus musculus
2	1488	76.0	380	11 Q9D695	Q9D695 mus musculus
3	800	40.8	388	11 Q91212	Q91212 mus musculus
4	798	40.7	388	11 Q9COV3	Q9COV3 mus musculus
5	763.5	39.0	423	11 Q9D7P9	Q9D7P9 mus musculus
6	708.5	36.2	379	11 Q9D154	Q9D154 mus musculus
7	707	36.1	382	11 Q8VHP7	Q8VHP7 mus musculus
8	700.5	35.8	379	11 Q8VHP7	Q8VHP7 mus musculus
9	694.5	35.5	369	4 Q9BYF7	Q9BYF7 mus musculus
10	686.5	35.0	387	11 Q9D105	Q9D105 mus musculus
11	684.5	34.9	377	11 Q9D804	Q9D804 mus musculus
12	684.5	34.9	385	11 Q8VHCO	Q8VHCO mus musculus
13	682	34.8	388	11 Q9Z2G2	Q9Z2G2 mus musculus
14	663	33.8	386	13 Q73860	Q73860 meleagris g
15	658	33.6	386	11 Q9D1E7	Q9D1E7 mus musculus
16	655	33.4	338	4 Q9BYF8	Q9BYF8 homo sapien

17	645	32.9	374	11 Q08797	Q08797 mus musculus
18	624.5	31.9	377	11 Q9DAV6	Q9DAV6 mus musculus
19	619	31.6	374	11 Q08800	Q08800 mus musculus
20	606.5	31.0	377	11 Q08806	Q08806 mus musculus
21	601.5	30.7	379	11 Q9D058	Q9D058 mus musculus
22	589.5	30.1	377	11 Q8VHQ1	Q8VHQ1 mus musculus
23	585.5	29.9	453	13 Q8VHQ2	Q8VHQ2 struthio ca
24	583.5	29.8	377	11 Q9DAZ7	Q9DAZ7 mus musculus
25	579.5	29.6	456	13 Q8VUS0	Q8VUS0 xenopus lae
26	573.5	29.3	453	13 Q8VUS1	Q8VUS1 chelydra se
27	570.5	29.1	452	13 Q9PFA8	Q9PFA8 salmo salar
28	570	29.0	426	17 Q8TNN7	Q8TNN7 methanosarc
29	568	29.0	410	13 Q73790	Q73790 gallus gall
30	565.5	28.9	423	13 Q91422	Q91422 gallus gall
31	564	28.8	455	4 Q13815	Q13815 homo sapien
32	555	28.3	359	11 Q9D6A7	Q9D6A7 mus musculus
33	553	28.2	459	11 Q9WMT1	Q9WMT1 cavia porce
34	550.5	28.1	424	17 Q8TKL5	Q8TKL5 methanosarc
35	542.5	27.7	448	13 Q9W648	Q9W648 fuqu rubrip
36	536.5	27.4	452	13 Q8VUX2	Q8VUX2 sphenodon p
37	535	27.3	377	5 Q9GPI3	Q9GPI3 ixodes ricl
38	510.5	26.1	380	5 Q8WDX0	Q8WDX0 thripicephal
39	496	25.3	334	12 Q9DHG4	Q9DHG4 yaba-like d
40	476.5	24.3	407	4 Q9UNU9	Q9UNU9 homo sapien
41	476	24.3	418	11 Q03734	Q03734 mus musculus
42	468.5	23.9	359	5 Q17365	Q17365 caenorhabdi
43	468.5	23.9	418	11 Q9D7D2	Q9D7D2 mus musculus
44	467.5	23.9	404	6 Q9N212	Q9N212 bos taurus
45	467	23.8	373	5 Q9NH65	Q9NH65 trichinella

## ALIGNMENTS

RESULT 1  
ID Q9D695 PRELIMINARY; PRT; 380 AA.  
AC Q9D695:  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE 4631416M05RIK protein (Megsin).  
GN SERPINB7 OR 4631416M05RIK.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=SKIN;  
RX MEDLINE=21085660; Pubmed=11217851;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamada I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kusakawa T., Saito R.,  
RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
RA Schirini L.M., Staubli F., Suzuki K., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carlini P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya K., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilmberg L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
RA Hayashizaki Y.,  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
RN [2]  
RP SEQUENCE FROM N.A.

```

RX MEDLINE-21368006; PubMed-11473647;
RA Nangaku M., Miyata T., Suzuki D., Umezono T., Hashimoto T., Wada T.,
RA Yagi M., Nagano N., Inagi R., Kurokawa K.;
RT "Cloning of rodent megalin revealed its up-regulation in
RT mesangio proliferative nephritis.";
RL Kidney Int. 60:641-652(2001).
DR EMBL; AK014524; BAB29410.1; -.
DR EMBL; AF105328; AAL16768.1; -.
DR HSSP; P05121; 1A7C.
DR MGD; MGI:2151053; Serpinb7.
DR InterPro; IPR000215; Serpin.
DR Pfam; PF00079; serpin; 1.
DR SMART; SM00093; SERPIN; 1.
DR PROSITE; PS00284; SERPIN; UNKNOWN_1.
SQ SEQUENCE 380 AA; 43050 MW; C9240272BCFB9CF4 CRC64;

Query Match 76.8%; Score 1505; DB 11; Length 380;
Best Local Similarity 73.9%; Pred. No. 1.2e-95;
Matches 281; Conservative 53; Mismatches 46; Indels 0; Gaps 0;

QY 1 MASLAANAEFCNLEFREMDDNGNGNVFESSLSLFAALALVRLGAQDDSLQIDKLLHV 60
DB 1 MASLAANAEFGDLEFREMDSOGNGNVFESSLSITFALTLLRLGARGCARQIDKALHF 60
QY 61 NTASGYGNSNSOGSLQSLKRVFSDINASHKDYLSYNGLFAEKVYGFHKDYIECAEK 120
DB 61 NTASGYGNSNSOGSLQSLKRVFSDINASHKDYLSYNGLFAEKVYGFHKDYIECAEK 120
QY 121 LVNAKVERDFTNHDTRNINIKWVNEETHGKIKKVIIEGGSSSAVAVLVNAVYFKK 180
DB 121 LVNAKVERDFTNVDQDTRKIKKWIENETHGKIKKVLGSSLSAVALNAVYFKK 180
QY 181 WQSAFTKSETINCHFKSPKCSKAVAMMHQERKFNLSVIEDPSMKITELRYNGIMNYVL 240
DB 181 WKSAFTKSDTLCSEFKSPGPKAVAMMHQERFNLSITQEPMPQVLELQYHGISMYYIM 240
QY 241 LPENDLSEIENKLTFFONLMEWTNPRRTSKYVEVFPPQRIEKNYEMKOYLRALGLKDI 300
DB 241 LPEADLSEIENKLTFFONLMEWTNPRRTSKYVEVFPPQRIEKNYEMKOYLRALGLKDI 300
QY 301 DESKALSGIASGRIYSRMHKSYLETEBETEAATGSIYVKOLPOSTLFRADHP 360
DB 301 DESKALSGIASGRIYSRMHKSYLETEBETEAATGSIYVKOLPOSTLFRADHP 360
QY 361 FLFVIRKDDIIFSGRVSCP 380
DB 361 FLFVIRKDDIIFSGRVSCP 380
QY 361 FLFVIRKDDIIFSGRVSCP 380
DB 361 FLFVIRKDDIIFSGRVSCP 380

RESULT 2
Q920J5 PRELIMINARY; PRT; 380 AA.
ID 0920J5;
AC 0920J5;
DT 01-DEC-2001 (TEMBLrel. 19, Created)
DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TEMBLrel. 20, Last annotation update)
DE Megsin.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-21368006; PubMed-11473647;
RA Nangaku M., Miyata T., Suzuki D., Umezono T., Hashimoto T., Wada T.,
RA Yagi M., Nagano N., Inagi R., Kurokawa K.;
RT "Cloning of rodent megalin revealed its up-regulation in
RT mesangio proliferative nephritis.";
RL Kidney Int. 60:641-652(2001).
RL -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.
DR EMBL; AF105329; AAL16769.1; -.
DR InterPro; IPR000215; Serpin.
DR Pfam; PF00079; serpin; 1.

```

```

DR PROSITE; PS00284; SERPIN; UNKNOWN_1.
KW Serpin.
SQ SEQUENCE 380 AA; 42821 MW; D8076CA8EE2C2PBC CRC64;

Query Match 76.0%; Score 1488; DB 11; Length 380;
Best Local Similarity 73.7%; Pred. No. 1.8e-94;
Matches 280; Conservative 57; Mismatches 43; Indels 0; Gaps 0;

QY 1 MASLAANAEFCNLEFREMDDNGNGNVFESSLSLFAALALVRLGAQDDSLQIDKLLHV 60
DB 1 MASLAANAEFGDLEFREMDSOGNGNVFESSLSITFALTLLRLGARGCARQIDKALHF 60
QY 61 NTASGYGNSNSOGSLQSLKRVFSDINASHKDYLSYNGLFAEKVYGFHKDYIECAEK 120
DB 61 NTASGYGNSNSOGSLQSLKRVFSDINASHKDYLSYNGLFAEKVYGFHKDYIECAEK 120
QY 121 LVNAKVERDFTNHDTRNINIKWVNEETHGKIKKVIIEGGSSSAVAVLVNAVYFKK 180
DB 121 LVNAKVERDFTNHDTRNINIKWVNEETHGKIKKVLGSSLSAVALNAVYFKK 180
QY 181 WQSAFTKSETINCHFKSPKCSKAVAMMHQERKFNLSVIEDPSMKITELRYNGIMNYVL 240
DB 181 WKSAFTKSDTLCSEFKSPGPKAVAMMHQERFNLSITQEPMPQVLELQYHGISMYYIM 240
QY 241 LPENDLSEIENKLTFFONLMEWTNPRRTSKYVEVFPPQRIEKNYEMKOYLRALGLKDI 300
DB 241 LPEADLSEIENKLTFFONLMEWTNPRRTSKYVEVFPPQRIEKNYEMKOYLRALGLKDI 300
QY 301 DESKALSGIASGRIYSRMHKSYLETEBETEAATGSIYVKOLPOSTLFRADHP 360
DB 301 VESRADLSGIASGRIYSRMHKSYLETEBETEAATGSIYVKOLPOSTLFRADHP 360
QY 361 FLFVIRKDDIIFSGRVSCP 380
DB 361 FLFVIRKDDIIFSGRVSCP 380

RESULT 3
Q91212 PRELIMINARY; PRT; 388 AA.
ID 091212;
AC 091212;
DT 01-DEC-2001 (TEMBLrel. 19, Created)
DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE RIKEN cDNA 2310046M08 gene.
GN 2310046M08RIK.
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RA Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.
DR EMBL; BC010313; AAH10313.1; -.
DR MGD; MGI:1914207; 2310046M08RIK.
DR InterPro; IPR000215; Serpin.
DR Pfam; PF00079; serpin; 1.
DR Serpin.
SQ SEQUENCE 388 AA; 43544 MW; AF4BB62FB084E6FA CRC64;

Query Match 40.8%; Score 800; DB 11; Length 388;
Best Local Similarity 41.8%; Pred. No. 3.6e-47;
Matches 162; Conservative 81; Mismatches 137; Indels 8; Gaps 5;

QY 1 MASLAANAEFCNLEFREMDDNGNGNVFESSLSLFAALALVRLGAQDDSLQIDKLLHV 60
DB 1 MDPITTAATPEFCADVKEKLSNNVGENIFESPPLTFYALSMILLGTRGKSABQMEKVLHY 60
QY 61 NTASGYGNSNSOGSLQSLKRVFSDINASHKDYLSYNGLFAEKVYGFHKDYIECAEK 120
DB 61 NTASGYGNSNSOGSLQSLKRVFSDINASHKDYLSYNGLFAEKVYGFHKDYIECAEK 120
QY 241 LPENDLSEIENKLTFFONLMEWTNPRRTSKYVEVFPPQRIEKNYEMKOYLRALGLKDI 300
DB 61 DSFSGVLKAKTRNSSECSQVGMHDPDFRALLISHINOONLSVANNRYGTGRSISFHKQYVR 120

```



KW Serpin. 423 AA; 47834 MW; 5A22BE2FE51B6120 CRC64;  
 SQ SEQUENCE

Query Match  
 Best Local Similarity 37.8%; Score 763.5; DB 11; Length 423;  
 Matches 160; Conservative 77; Mismatches 143; Indels 43; Gaps 4;

QY 1 MASLAANAEFCENLREMDNDGNGNVPFSSLSLPAALALVRLGADDSLSQIDKLLHV 60  
 1 MDLSLTAAANKFCDFPREISKDDAHKNIIFVCPILSLAFAFGWRLGARQSDAHQIDEALH 60  
 DB 61 NTAS-----GYNSSNSOSGLQSLKRVFSD 86  
 DB 61 NELSKDEHKEPNDPSPQSEKASDSSLEGGKQTSASODQOGEESTNHOLGCHFGKLLSR 120  
 QY 87 INASHRDYDLSIYNGLFAEKVYGFHNDYIECAEKLVDKADYERDFTNHLEDFRRNINRW 146  
 121 IDRDKSYITLTMANKRLYGDEDFICSEYSDVTEFHHTVESYDFQKDSKSHQELNFW 180  
 DB 147 EMTHTGKIRNVIGEGISSAVVLYNAVYFKGKMSAFTSEITNCHFKSPKSGKAVA 206  
 181 ESGSGKIKELFEKALINDSTVLVNAVYFKAKWREFNSEMTVDSFCLNENEKITY 240  
 QY 207 MMHQEKFNLSYEDPSMKILELY-NGGIMTVLLPE-----NDLSEIENKLFYQNM 259  
 DB 241 MNMOKGKFRIGFIDELQAOILKMKYAMGKLSMLVLLPSCSESDVNSLOELEKINHEKLL 300  
 QY 260 EMTNPRMTSKYVEVFPQKIEKNYEMKOYLRLGLKIDFDRKADLSGIASGRLYIS 319  
 DB 301 ANSSSENLSKPRPAIFPOPNLESDYDLSIIDDMKIKDVFETKADLGLGISPNVLYIS 360  
 QY 320 RMWHSYIEVTEGETATAGTSNIVEKOLPOSTLFRADHPLELVIRKD--DILFSGKY 377  
 DB 361 KIYHKFFVEVDEMGTOAAAGVAAEKALPFWVEFNANHPFLFIRHNPQTGLRCGRV 420  
 QY 378 SCP 380  
 DB 421 YCP 423

RESULT 6  
 ID 09D154 PRELIMINARY; PRT; 379 AA.  
 AC 09D154;  
 DT 01-JUN-2001 (Tremblrel. 17, Created)  
 DT 01-JUN-2001 (Tremblrel. 17, last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, last annotation update)  
 DE 1190005M04Rik protein (RIKEN CDNA 1190005M04 gene) (EIA).  
 GN SERPINB1A OR 1190005M04Rik OR SERPINB1.  
 OS Mus musculus (Mouse).  
 OC Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.  
 OX NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=EMBRYO;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka T.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Glasi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schiml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momhaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyo-Oka K., Wang K.H., Weitz C., Whittaker C., Wilmink L.,

RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,  
 RA Hayashizaki Y.,  
 RT "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=COLON;  
 RA Strausberg R.;  
 RN Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=SPLEEN;  
 RA Kaiseerman D., Knaggs S., Scarif K.L., Gillard A., Mirza G., Cadman M.,  
 RA McKee R., Denny P., Cooley J., Benarafa C., Remold-O'Donnell E.,  
 RA Ragousis J., Bird P.I.;  
 RT "Comparison of human chromosome 6p25 with murine chromosome 13 reveals  
 a greatly expanded O<sub>2</sub>-serpin gene repertoire in the mouse."  
 RL Genomics 0:0-0(2002).  
 CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.  
 DR EMBL: AK003930; BAB23079.1; -;  
 DR EMBL: BC011140; AAH11440.1; -;  
 DR EMBL: AF426024; AAL57486.1; -;  
 DR HSP: P05619; IHL.  
 DR MGD: MGI:1913472; Serpinb1a.  
 DR InterPro: IPR000215; Serpin.  
 DR Pfam: PF00079; Serpin.  
 DR SMART: SM00093; Serpin.1.  
 DR PROSITE: PS00284; SERPIN.1.  
 KW Serpin.  
 SQ SEQUENCE 379 AA; 42575 MW; F492F1EBB8E80B1 CRC64;  
 Query Match  
 Best Local Similarity 36.2%; Score 708.5; DB 11; Length 379;  
 Matches 145; Conservative 81; Mismatches 144; Indels 19; Gaps 4;

QY 1 MASLAANAEFCENLREMDNDGNGNVPFSSLSLPAALALVRLGADDSLSQIDKLLHV 60  
 1 MDLSLTAAANKFCDFPREISKDDAHKNIIFVCPILSLAFAFGWRLGARQSDAHQIDEALH 60  
 DB 61 NTASGYNSNSOSGLQSLKRVFSDINASHRDYDLSIYNGLFAEKVYGFHNDYIECAEKL 110  
 61 -----DSVEDIHSRFOSSINAEVSRGASHITKLNRIGERTYFLEPYLASTOK 110  
 QY 121 LYDAKAVYDFNTHLEDFRRNINRWENTHCKIKKAVIGEGISSAVVLYNAVYFKGK 180  
 111 MGADLAPVDPLHASDARKELNQWYKGTGEPILSLVGVVDSMTKVLVNAVYFKGM 170  
 QY 181 WSAFTKSETINCHFKSPKSGKAVAMMHQERKFNLSYEDPSMKILELYNGG-1NMYV 239  
 171 WEKFEWTEDTTDAPEFLSKKDKTYKMMYOKKKFPFGYISDLCKVLEMPYOGGELSMVI 230  
 QY 240 LIPEN-----DLSEIENKLFYQNMEMWTPNPRMTSKYVEVFPQKIEKNYEMKOYLRA 293  
 DB 231 LILPKDIEDESGTLKIEKQITLLEMTKRENLEFDVHVLPRKIEESTLNSNLR 290  
 QY 294 LGLKIDFESKADLSGIASGRLYISRMWHSYIEVTEGETATAGTSNIVEKOLPOST 353  
 291 LGVQDLFSSKADLSGMSGRDLFLISKIYHKFFVEVNEGSTAATAATGIAIFCMILPBE 350  
 QY 354 LFRADHPLELVIRKDDI--ILFSGKYSCP 380  
 DB 351 EFTVDHPLEFIRHNPSTNVLELGRVCSF 379

RESULT 7  
 ID 08VHP7 PRELIMINARY; PRT; 382 AA.  
 AC 08VHP7;  
 DT 01-MAR-2002 (Tremblrel. 20, Created)  
 DT 01-MAR-2002 (Tremblrel. 20, last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, last annotation update)  
 DE EIB.  
 GN SERPINB1B.



OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BAB/C; TISSUE=LUNG;  
 RA Kalsnerman D., Knaggs S., Scarff K.L., Gillard A., Mirza G., Cadman M.,  
 RA McKone R., Denny P., Cooley J., Benarafa C., Remold-O'Donnell E.,  
 RA Ragusa J., Bird P.I.;  
 RT "Comparison of human chromosome 6p25 with murine chromosome 13 reveals  
 a greatly expanded *Ov-serpin* gene repertoire in the mouse."  
 RL Genomics 0:0-0(2002).  
 DR EMBL: AF426025; AAL57487.1;  
 DR InterPro: IPR002016; Peroxidase.  
 DR InterPro: IPR00215; Serpin.  
 DR Pfam: PF00079; serpin.1.  
 DR SMART: SM00093; SERPIN.1.  
 DR PROSITE: PS00435; PEROXIDASE.1; UNKNOWN.1.  
 DR PROSITE: PS00284; SERPIN; UNKNOWN.1.  
 SQ SEQUENCE 382 AA; 42887 MW; 35CB6AD677C8DB CRC64;

Query Match 36.1%; Score 707; DB 11; Length 382;  
 Best Local Similarity 38.8%; Pred. No. 8.7e-41;  
 Matches 152; Conservative 76; Mismatches 142; Indels 22; Gaps 7;

OY 1 MASLAANAECFENLFREMDNOGNGNFFSSLSFALALVRLGAODSLSQDKLHV 60  
 DB 1 MEOLSSANTLTLELFHTLKESPFGNIFFPFSSISLAVFLGAKGSTAOLSKTFLH- 59  
 OY 61 NTASGYGNSNSQSGLSQSLKRVFSDINASHKDYDLSYNGLFARVYGFHNDYIECAEK 120  
 DB 60 -----FDSVEDIHSCFQSLTALEV-SKLGASH--TLKLANRLYGEKTYNFLDEPLASQK 110  
 OY 121 LYDAVERVDFTNHLIEDTRRNKKNVENETHGKIRNVGEGISSAVAVLVNAYYFGK 180  
 DB 111 MSADLAADVQHAESDARKEINQWVKQTEGKIPELLAKGVADSMTKLVLNATYFGKI 170  
 OY 181 WOSATKSETINCHFKSPKCSGKAVAMHOKERKFLSVIEDPSMKILELRYNGG-INMY 239  
 DB 171 WEQMTRETINAPRLNKKDKTKYKMMYQKKKFFPGYISDLCKCVLEMPYOGGELSMT 230  
 OY 240 LLPEN-----DLSEIENKLTFFONLMEWTNPRRMTSKYVEVFPQFKIEKNEMQYLRA 293  
 DB 231 LLPEDIEDESGTLKIEQDLTGKLEHMTKHENLNIDVHVLPFRKMEESTYLSNLC 290  
 OY 294 LGLKIDFESKADLSGIASGRLYISRMHKSIEVTEBGEATATGSGNT--VEKOLP 350  
 DB 291 LGVODLFSSKADLSGMSGSRDLFVSKIVHKSFDVNEGQTEAAATGIIQVLECKMPT 350  
 OY 351 OSTFRADHPFLFVIRKDDI--ILFSKGVSCP 380  
 DB 351 POFVTVDPHPLFIRHNPTANMIFGRVCS 382

RESULT 8  
 O9D7S8 PRELIMINARY: PRT: 379 AA.  
 AC O9D7S8;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE 1190005M04RIK protein.  
 GN SERPINB1 OR 1190005M04RIK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=STOMACH;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shimagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,

RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 RA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nkaido I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Stabili F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Balarelli R., Barsh G.,  
 RA Blake J., Boffelli P., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamliya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,  
 RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).  
 CC -1 SIMILARITY: BELONGS TO THE SERPIN FAMILY.  
 DR EMBL: AK008914; BAB25964.1;  
 DR HSSP: P05619; 1HLE.  
 DR MGD: MGI:1913472; Serpinb1a.  
 DR InterPro: IPR00215; Serpin.  
 DR Pfam: PF00079; serpin.1.  
 DR SMART: SM00093; SERPIN.1.  
 DR PROSITE: PS00284; SERPIN.1.  
 KW Serpin.  
 SQ SEQUENCE 379 AA; 42580 MW; 1937285850777B4B CRC64;

Query Match 35.8%; Score 700.5; DB 11; Length 379;  
 Best Local Similarity 38.6%; Pred. No. 2.4e-40;  
 Matches 150; Conservative 75; Mismatches 145; Indels 19; Gaps 6;

OY 1 MASLAANAECFENLFREMDNOGNGNFFSSLSFALALVRLGAODSLSQDKLHV 60  
 DB 1 MEOLSSANTLTLELFHTLKESPFGNIFFPFSSISLAVFLGAKGSTAOLSKTFLH- 59  
 OY 61 NTASGYGNSNSQSGLSQSLKRVFSDINASHKDYDLSYNGLFARVYGFHNDYIECAEK 120  
 DB 60 -----FDSVEDIHSPQSONAEV-SKRGASH--TLKLANRLYGEKTYNFLDEPLASQK 110  
 OY 121 LYDAVERVDFTNHLIEDTRRNKKNVENETHGKIRNVGEGISSAVAVLVNAYYFGK 180  
 DB 111 MYGADLAPDELHAESDARKEINQWVKQTEGKIPELLASGVADSMTKLVLNATYFGM 170  
 OY 181 WOSATKSETINCHFKSPKCSGKAVAMHOKERKFLSVIEDPSMKILELRYNGG-INMY 239  
 DB 171 WEKFMTEEDTDPAPRLSKDKTRTKYKMMYQKKKFFPGYISDLCKCVLEMPYOGGELSMT 230  
 OY 240 LLPEN-----DLSEIENKLTFFONLMEWTNPRRMTSKYVEVFPQFKIEKNEMQYLRA 293  
 DB 231 LLPKIDIEDESGTLKIEQDLTGKLEHMTKHENLNIDVHVLPFRKMEESTYLSNLCGR 290  
 OY 294 LGLKIDFESKADLSGIASGRLYISRMHKSIEVTEBGEATATGSGNTYIEKOLPOST 353  
 DB 291 LGVODLFSSKADLSGMSGSRDLFVSKIVHKSFDVNEGQTEAAATGIIQVLECKMPT 350  
 OY 354 LFRADHPFLFVIRKDDI--ILFSKGVSCP 380  
 DB 351 EFTVDPHPLFIRHNPTSNVLELGRVCS 379

RESULT 9  
 O9BYF7 PRELIMINARY: PRT: 369 AA.  
 AC O9BYF7;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE SCCA2b.  
 GN SCCA2.  
 OS Homo sapiens (Human).

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Sunami Y., Kishi F., Murakami A., Sakaguchi Y., Kato H.;
RT "Novel forms of SCC Antigen transcripts produced by alternative
RT splicing.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.
DR EMBL: AB046400; BAB40773.1;
DR HSSP: P01008; IATH.
DR InterPro: IPR000215; Serpin.
DR Pfam: PF00079; serpin; 1.
DR SMART: SM00093; SERPIN; 1.
DR PROSITE: PS00284; SERPIN; 1.
KW Serpin.
SQ
SEQUENCE 369 AA; 42287 MW; CHD7A2426378DE3E CRC64;

Query Match 35.5%; Score 694.5; DB 4; Length 369;
Best Local Similarity 39.0%; Pred. No. 6e-40;
Matches 153; Conservative 82; Mismatches 122; Indels 35; Gaps 10;

QY 1 MASIAANAEFCFLFREMDNONGNNGVFFSSLSLFAALALVRLGADDSLSQIDKLHV 60
D 1 MNSLSEANTKFMFLFOOFRKSKEN-NIFYSPISTALGAVLLGARDNDAQOISKVLHF 59
DB
QY 61 -----NTASGYGNSNSQSG-IQSOLKRVFSDINASHKDDLSIVNGLFAEKVGFHKDY 114
D 60 DQYVENTEKAATYHYDRSGVHHQFOKLTEFKSKSDAHELKANKLFGKTYQFOLEY 119
QY 115 IECAEKIYDAKVENVDFTNLEDRINKWENETGKIKNVIGEGISSAVVAVLVA 174
D 120 LDKATKRYQSVSESTDANAPESRKRKNISWVESQTEKIKNLEPPDGTIGDITLVVNA 179
QY 175 VYFCKGQSAFTKSETINCHFKSPKCGKAVAMHQRKFLSYIEDPSMKILELRNG- 233
D 180 IYFGQWENKFKKNT-----KEEKFPN--KVQAQVLEIPKKG 218
QY 234 GINMYVLLPE--NDLSEIENKLFQNLMEWTNPRRTSKYVEFFPOFKIEKNEMQYL 291
D 219 DLSMIVLLPMEIDQLKLEKLTAEKLMEMTSLQNMETGCVDDLPRFKMESDLDKDTL 278
QY 292 RALGLKIFDESKADLSGASGRLYISRMHKSIVTEBGEATAGSNIVKOLPQ 351
D 279 RINGMVAIFN-GADLDSGMTWSHLSVSKLHKAFAVEVEGVAATAATVAVVELSSP 337
QY 352 ST-LFRADHPFLFYIR--KDDILFSGKVSQP 380
D 338 TNEFCCHPFLPFIKONKTNISLTYGRFSSP 369
DB

RESULT 10
Q9D105 PRELIMINARY; PRT; 387 AA.
AC Q9D105;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE 18 days embryo cDNA, RIKEN full-length enriched library,
DE clone:1110001H02, full insert sequence.
GN SCCA2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX MEDLINE=21085660; Pubmed=11217851;
RA Kawai J., Shinaawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Kondo H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,

```

```

RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carlini P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Welter C., Whitaker C., Wilming L.,
RA Wyszawski-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.
DR EMBL: AK003220; BAB2650.1;
DR HSSP: P05120; Iby7.
DR MGD: MGI:127952; Scca2.
DR InterPro: IPR000215; Serpin.
DR Pfam: PF00079; serpin; 1.
DR SMART: SM00093; SERPIN; 1.
DR PROSITE: PS00284; SERPIN; 1.
KW Serpin.
SQ
SEQUENCE 387 AA; 44558 MW; AE25C02520F3B194 CRC64;

Query Match 35.0%; Score 686.5; DB 11; Length 387;
Best Local Similarity 36.8%; Pred. No. 2.3e-39;
Matches 143; Conservative 91; Mismatches 144; Indels 11; Gaps 6;

QY 1 MASIAANAEFCFLFREMDNONGNNGVFFSSLSLFAALALVRLGADDSLSQIDKLHV 59
D 1 MIRHADVAFVAVEMTQQL--RESKNIIFSPISMALMLQAGAGNEIQEVLQF 58
QY 60 VNTASGYGNS--NSQSGIQSOLKRVFSDINASHKDDLSIVNGLFAEKVGFHKDY 116
D 59 IETTKTKTESSECHDDDEENHEQFLITOLNNSNDYDKAANSIYGAGFPFLQFLE 118
QY 117 CAEKLYDAKVENVDFTNLEDRINKWENETGKIKNVIGEGISSAVVAVLVA 176
D 119 DIKEYQAKVESLDFEATSESEKIKNSWESKINGIKDLPSSGLSSSTILVAVNY 178
QY 177 FKGQWASAFKSETINCHFKSPKCGKAVAMHQRKFLSYIEDPSMKILELRNG-GI 235
D 179 FKGQWKNKFMENHTRREKFLNKNKTSKPVQMMQRNKNFNSPLGDVAQAVLPIYKQDL 238
QY 236 NMVYVLLPE--NDLSEIENKLFQNLMEWTNPRRTSKYVEFFPOFKIEKNEMQYLRA 293
D 239 SMEVLLPMEIDQLKLEKLTAEKLMEMTSLQNMETGCVDDLPRFKMESDLDKDTL 278
QY 294 LGLKIDFDESKADLSGASGRLYISRMHKSIVTEBGEATAGSNIVKOLPQ 353
D 299 MGVADADPQKADSGMSLIPGLVSKVLKSFYVNEBGEATAAGVEVSVRAQIAE 358
QY 354 LFRADHPFLFYIR--RKDDILFSGKVSQP 380
D 359 DFCDDHPFLFYIRHMNINSLTYGRFSSP 387
DB

RESULT 11
Q008804 PRELIMINARY; PRT; 377 AA.
AC Q008804;
DT 01-JUN-1997 (TREMBlrel. 04, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Serine proteinase inhibitor NK13.
GN SP12.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL;  
 RX MEDLINE=9736124; PubMed=9182575;  
 RA Sun J., Ooms L., Bird C.H., Sutton V.R., Trapani J.A., Bird P.I.;  
 RT "A new family of 10 murine ovalbumin serpins includes two homologs of  
 RT proteinase inhibitor 8 and two homologs of the granzyme B inhibitor  
 RT (proteinase inhibitor 9).";  
 RL J. Biol. Chem. 272:15434-15441(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL;  
 RA Kalseman D., Knaggs S., Scarff K.L., Gillard A., Mirza G., Cadman M.,  
 RA McKee R., Denny P., Cooley J., Benarafa C., Remold-O'Donnell E.,  
 RA Ragousis J., Bird P.I.;  
 RT "Comparison of human chromosome 6p25 with murine chromosome 13 reveals  
 RT a greatly expanded Ov-serpin gene repertoire in the mouse.";  
 RL Submitted (SEP-2001) to the EMBL/Genbank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.  
 DR EMBL: U96707; AAB57819.2; -;  
 DR MGD: MGI:894688; SP112;  
 DR InterPro: IPR000215; Serpin.  
 DR Pfam: PF00079; serpin; 1.  
 DR SMART: SM00093; SERPIN; 1.  
 DR PROSITE: PS00284; SERPIN; 1.  
 KW Serpin.  
 SQ SEQUENCE 377 AA; 42535 MW; 96608E8F3A41EE47A CRC64;  
 Query Match 34.9%; Score 684.5; DB 11; Length 377;  
 Best Local Similarity 39.4%; Pred. No. 3e-39;  
 Matches 152; Conservative 76; Mismatches 143; Indels 15; Gaps 8;  
 OY 1 MASLAANAEEFCNLFREMDNNGNNGVFFSSLSLFAALALVRLGADDSLSQIDKLHV 60  
 DB 1 MDPLLEANAATFALNLKLTIGEDSSR-NVFLSPISISSALVAVFGAGTATASQAQALSL 59  
 OY 61 NTASGYGSSNSGSG-LQSOLKRVFSDINASHKDYDLSYNGLEFAEYVGFHRYICAE 120  
 DB 60 DKSGGKGRDVH-QFOSL---TETNKTGTQVLTANLFEKFTDILASKDCRK 114  
 OY 121 LYDAKVERVDFTNHLDTNRNINKWENETHGKIKNVIGEGISSAVMLVNAVYFEGK 180  
 DB 115 FYEAEEMELDFKATQDSQHINAWAKTEDEKTELSSGVSNSNPLVNAVYFEGK 174  
 OY 181 WQSAFTKSETINCFKSPKSGKAVAMHDERKNSLYIEDPSKILRLYNGG-INMYV 239  
 DB 175 WEKQFNKEDQEMFNTKDVVKPVQMFQKSTFKMYVEEISTNILLPYVGNELMMII 234  
 OY 240 LLEPEN--DLSEIENKLTFOMLMEWNTNPRMTSKYVEVFPQFKTEKVENKQYLRALGK 297  
 DB 235 MLPEHIELSNVEKELTYKFIETRLDKMEEEVEVFLPREFKLEENVDKMDVLCRLGNT 294  
 OY 298 DIFDESKADLSGIASGRLYISRMHKSYLEVEEGTEATATGNSNI-VEKQLPOSTLFR 356  
 DB 295 DAFEEGNADSGIASKRGFLSKYIHKSIFYEVNEEGTEAATAATANIGFRCMV---YFC 351  
 OY 357 ADHPFLFVIR--KDDIILFSGKVS 380  
 DB 352 ANHPFLFIQHSRTSGIVFCGRFSSP 377  
 RESULT 12  
 ID 08VH00 PRELIMINARY; PRT; 385 AA.  
 AC 08VH00;  
 DT 01-MAR-2002 (TReMBLrel. 20, Created)  
 DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE SP13L2.  
 GN SP13L2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B6D F1/J;  
 RA Kalseman D., Knaggs S., Scarff K.L., Gillard A., Mirza G., Cadman M.,  
 RA McKee R., Denny P., Cooley J., Benarafa C., Remold-O'Donnell E.,  
 RA Ragousis J., Bird P.I.;  
 RT "Comparison of human chromosome 6p25 with murine chromosome 13 reveals  
 RT a greatly expanded Ov-serpin gene repertoire in the mouse.";  
 RL Submitted (SEP-2001) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AF425084; AAL65910.1; -;  
 DR InterPro: IPR000215; Serpin.  
 DR Pfam: PF00079; serpin; 1.  
 DR SMART: SM00093; SERPIN; 1.  
 DR PROSITE: PS00284; SERPIN; UNKNOWN\_1.  
 SQ SEQUENCE 385 AA; 43786 MW; 0820A8A26AEBB485 CRC64;  
 Query Match 34.9%; Score 684.5; DB 11; Length 385;  
 Best Local Similarity 39.6%; Pred. No. 3.1e-39;  
 Matches 156; Conservative 75; Mismatches 140; Indels 23; Gaps 9;  
 OY 1 MASLAANAEEFCNLFREMDNNGNNGVFFSSLSLFAALALVRLGADDSLSQIDKLHV 60  
 DB 1 MDPLLEANAATFALNLKLTIGEDRSK-NVFLSPISISSALVAVFGAGTATATQALSL 59  
 OY 61 NTASGYGSSNSGSG-LQSOLKRVFSDINASHKDYDLSYNGLEFAEYVGFHRYICAE 119  
 DB 60 -----GRCSSESDGDVHOGFQLLSSEVNTKGTQYSLKANRLGEKFTDILASFQSC 113  
 OY 120 KLYDAKVERVDFTNHLDTNRNINKWENETHGKIKNVIGEGISSAVMLVNAVYFEGK 179  
 DB 114 FYEAEEMELDFKATQDSQHINAWAKTEDEKTELSSGVSNSNPLVNAVYFEGK 173  
 OY 180 WQSAFTKSETINCFKSPKSG-----KAVAMHDERKNSLYIEDPSKILRLYNGG 232  
 DB 174 WEKQFNKEDQEMFNTKDVVKPVQMFQKSTFKMYVEEISTNILLPYV 233  
 OY 233 GG-INMYVLEPEN--DLSEIENKLTFOMLMEWNTNPRMTSKYVEVFPQFKTEKVENKQ 289  
 DB 234 GNEINMIMLPDEHVELSTYEKEITHEKFTETRLDRKGGKRVFLPWFLEBNYMKD 293  
 OY 290 YLRALGKIDIFDESKADLSGIASGRLYISRMHKSYLEVEEGTEATATGNSNIYEQ 349  
 DB 294 VLCKLGMIDAFEEGRADPSGSSKQGLFLSVIHKSIFYEVNEEGTEAATAAT--TIVLKS 351  
 OY 350 POST-LPRADHPFLFVIR--KDDIILFSGKVS 380  
 DB 352 SRSTPCFCVNRPFIFIOHKTNETILFLGRSSP 385  
 RESULT 13  
 ID 0922G2 PRELIMINARY; PRT; 388 AA.  
 AC 0922G2;  
 DT 01-MAY-1999 (TReMBLrel. 10, Created)  
 DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
 DE Squamous cell carcinoma antigen 2.  
 GN SCCA2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=99047536; PubMed=9828132;  
 RA Bartuski A.J., Kamachi Y., Schlick C., Massa H., Trask B.J.,  
 RA Silverman G.A.;  
 RT "A murine ortholog of the human serpin SCGA2 maps to chromosome 1 and  
 RT inhibits chymotrypsin-like serine proteinases.";  
 RL Genomics 54:297-306(1998).  
 CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.



DR Pfam; PF00079; serpin; 1.  
DR SMART; SM00093; SERPIN; 1.  
DR PROSITE; PS00284; SERPIN; 1.  
KW Serpin.  
SQ SEQUENCE 386 AA; 44313 MW; 2A3CFA78A73AD004 CRC64;

Query Match 33.6%; Score 658; DB 11; Length 386;  
Best Local Similarity 36.0%; Pred. No. 2,1e-37;  
Matches 138; Conservative 92; Mismatches 141; Indels 12; Gaps 6;

```
OY 7 ANAEFCNLPREMDNNGNGVFPSSLSLPAALALVRLGAODDSLQIDKLHVNTASGY 66
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 7 ADGKFTVEMYROL--RESDKNIFFSPISMITALGMLKLGAKGNTIIEKYLQCNETTEK 64
OY 67 GNSNS----QSGLSQLKRVFSDINASHRDYDLSIVNGLEPAEKVYGFHKDYTECAEKLY 122
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 65 TTEKSAHCDDNDNVHEQFOKLITQLNKSNDYDLKANSIYGAKGFPLQLTFLEDIKEY 124
OY 123 DAKYRVDFTNHLEDTRRNINKVVENETHGKIKNVIGEGISSAVNVLVNAVYFKGKMQ 182
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 125 HANVESIDFEHAAEESSEKKNFSYKNETNGKIKDLFPSSGLSSSTKLVLVNAVYFKGRWN 184
OY 183 SAFTKSETINCHFKSPKCGKAVAMHQAERFNLSTIEDPSMKITLRLYNG-GIMNVLL 241
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 185 HKFDENNTIEEMFVLNKNNTSIPVPMKQRNKFMSFLEDVQAQIVEIPIYKKELSMFVLR 244
OY 242 PE--NDLSEIENKLTTPQNLMEWTNPRRMTSKYVEVFFPOFKIEKNYEMKOYLRLALGKDI 299
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 245 PMEIDGLKQLEKQLTAAKLEWTRAEHMHLELYLWLPREFVEKEKYDLPVPLECGMGNVA 304
OY 300 FDESKADLSCIASGRLYISRMHHSYIEVTEGTEATAATGSNIVEKQLPOSTLFRADH 359
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 305 FDPQKADFGSGMSSSTQGLVGVKVLHKSFEVNEEGTEADPASGEVY-LRLAQVADFRCDH 363
OY 360 PFLFVI--RKDDIILFSGKVSCP 380
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 364 PFLFFIHSKTSILFFGRISSP 386
```

Search completed: July 11, 2003, 11:59:12  
Job time : 167.904 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:54:32 ; Search time 28.3285 Seconds

(Without alignments)  
204.833 Million cell updates/sec

Title: US-10-091-442-1

Perfect score: 52

Sequence: 1 SETINCHFK 9

Scoring table: BLOSUM62

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending\_Patents\_AA\_Main:\*

1: /cgn2\_6/ptodata/1/paa/PCPUS\_COMB.pep.\*

2: /cgn2\_6/ptodata/1/paa/US05\_COMB.pep.\*

3: /cgn2\_6/ptodata/1/paa/US07\_COMB.pep.\*

4: /cgn2\_6/ptodata/1/paa/US08\_COMB.pep.\*

5: /cgn2\_6/ptodata/1/paa/US081\_COMB.pep.\*

6: /cgn2\_6/ptodata/1/paa/US082\_COMB.pep.\*

7: /cgn2\_6/ptodata/1/paa/US083\_COMB.pep.\*

8: /cgn2\_6/ptodata/1/paa/US084\_COMB.pep.\*

9: /cgn2\_6/ptodata/1/paa/US085\_COMB.pep.\*

10: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep.\*

11: /cgn2\_6/ptodata/1/paa/US087\_COMB.pep.\*

12: /cgn2\_6/ptodata/1/paa/US088\_COMB.pep.\*

13: /cgn2\_6/ptodata/1/paa/US09\_COMB.pep.\*

14: /cgn2\_6/ptodata/1/paa/US090\_COMB.pep.\*

15: /cgn2\_6/ptodata/1/paa/US091\_COMB.pep.\*

16: /cgn2\_6/ptodata/1/paa/US092\_COMB.pep.\*

17: /cgn2\_6/ptodata/1/paa/US093\_COMB.pep.\*

18: /cgn2\_6/ptodata/1/paa/US094\_COMB.pep.\*

19: /cgn2\_6/ptodata/1/paa/US095\_COMB.pep.\*

20: /cgn2\_6/ptodata/1/paa/US096\_COMB.pep.\*

21: /cgn2\_6/ptodata/1/paa/US097\_COMB.pep.\*

22: /cgn2\_6/ptodata/1/paa/US098\_COMB.pep.\*

23: /cgn2\_6/ptodata/1/paa/US099\_COMB.pep.\*

24: /cgn2\_6/ptodata/1/paa/US100\_COMB.pep.\*

25: /cgn2\_6/ptodata/1/paa/US101\_COMB.pep.\*

26: /cgn2\_6/ptodata/1/paa/US102\_COMB.pep.\*

27: /cgn2\_6/ptodata/1/paa/US60\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query length	ID	Description
1	52	100.0	9	US-09-140-719-1
2	52	100.0	9	US-10-091-442-1
3	52	100.0	380	US-08-091-028A-34
4	52	100.0	380	US-09-140-719-34
5	52	100.0	380	US-09-508-997A-2
6	52	100.0	380	US-09-791-537-281

7	52	100.0	380	US-09-936-883A-2	Sequence 2, Appl1
8	52	100.0	380	US-10-091-442-34	Sequence 34, Appl1
9	52	100.0	380	US-10-110-030-2	Sequence 2, Appl1
10	39	75.0	23	US-09-724-059-151732	Sequence 151732,
11	39	75.0	23	US-09-724-059-154252	Sequence 154252,
12	39	75.0	23	US-09-724-059-156652	Sequence 156652,
13	39	75.0	23	US-09-724-059-159172	Sequence 159172,
14	39	75.0	23	US-09-724-059-161692	Sequence 161692,
15	39	75.0	23	US-09-724-059-163972	Sequence 163972,
16	39	75.0	23	US-09-724-059-166732	Sequence 166732,
17	39	75.0	23	US-09-724-059-169012	Sequence 169012,
18	39	75.0	23	US-09-724-059-171412	Sequence 171412,
19	39	75.0	23	US-09-724-059-173932	Sequence 173932,
20	39	75.0	185	US-10-179-131-7804	Sequence 14712, A
21	39	75.0	185	US-10-179-131-7804	Sequence 7804, Ap
22	39	75.0	185	US-09-508-997A-4	Sequence 14712, A
23	39	75.0	380	US-09-508-997A-4	Sequence 4, Appl1
24	39	75.0	380	US-09-936-883A-19	Sequence 19, Appl1
25	39	75.0	380	US-10-110-030-4	Sequence 13518, A
26	38	73.1	199	US-09-614-150-13518	Sequence 13518, A
27	38	73.1	199	US-09-791-537-134773	Sequence 134773,
28	38	73.1	199	US-09-167-245-514	Sequence 514, App
29	38	73.1	199	US-09-167-245-514	Sequence 13559, A
30	38	73.1	199	US-09-191-637-13559	Sequence 10688, A
31	38	73.1	208	US-09-614-150-9252	Sequence 126235,
32	38	73.1	208	US-09-791-537-126235	Sequence 9252, Ap
33	38	73.1	208	US-09-167-245-386	Sequence 386, App
34	38	73.1	208	US-09-167-245-386	Sequence 7495, Ap
35	38	73.1	208	US-09-191-637-9281	Sequence 9281, Ap
36	38	73.1	208	US-09-191-637-9281	Sequence 7241, Ap
37	37	71.2	9	US-08-091-028A-1	Sequence 1, Appl1
38	37	71.2	9	US-08-091-028A-1	Sequence 1, Appl1
39	37	71.2	320	US-08-070-669A-5	Sequence 5, Appl1
40	37	71.2	320	US-08-172-428A-13	Sequence 13, Appl1
41	37	71.2	320	US-08-384-032-4	Sequence 4, Appl1
42	37	71.2	638	US-10-219-999-59911	Sequence 59911, A
43	37	71.2	1030	PCT-US01-29288-140	Sequence 140, App
44	36	69.2	23	US-09-724-059-151372	Sequence 151372,
45	36	69.2	23	US-09-724-059-152932	Sequence 152932,

#### ALIGNMENTS

RESULT 1

US-09-140-719-1

Sequence 1, Application US/09140719

GENERAL INFORMATION:

APPLICANT: TSUJIMOTO, Masafumi

APPLICANT: IMASA, Fuyuki

APPLICANT: TSUROOKA, Nobuo

APPLICANT: NAKAZATO, Hiroshi

APPLICANT: MIURA, Kenju

APPLICANT: ISHIDA, Nobuhiro

APPLICANT: KURIHARA, Tatsuya

APPLICANT: YAMACHI, Kozo

APPLICANT: YAMAGUCHI, Nozomi

TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR

NUMBER OF SEQUENCES: 34

CORRESPONDENCE ADDRESS:

ADDRESSEE: Burns, Doane, Swecker & Mathis

STREET: P.O. Box 1404

CITY: Alexandria

STATE: Virginia

COUNTRY: United States

ZIP: 22313-1404

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/140,719

FILING DATE: 08-AUG-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-6620  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-140-719-1

Query Match 100.0%; Score 52; DB 15; Length 9;  
Best Local Similarity 100.0%; Pred. No. 4.2e+06;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SETINCHK 9  
Db 1 SETINCHK 9

RESULT 2  
US-10-091-442-1  
Sequence 1, Application US/10091442  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
IMASA, Fuyuki  
TSURUOKA, Nobuo  
NAKAZATO, Hiroshi  
MIURA, Kenju  
ISHIDA, Nobuhiko  
KURIHARA, Tatsuya  
YAMAICHI, Kozo  
YAMAGUCHI, Nozomi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/091,442  
FILING DATE: 07-Mar-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998

APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-6620  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-10-091-442-1

Query Match 100.0%; Score 52; DB 24; Length 9;  
Best Local Similarity 100.0%; Pred. No. 4.2e+06;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SETINCHK 9  
Db 1 SETINCHK 9

RESULT 3  
US-08-091-028A-34  
Sequence 34, Application US/08091028A  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
IMASA, Fuyuki  
TSURUOKA, Nobuo  
NAKAZATO, Hiroshi  
MIURA, Kenju  
ISHIDA, Nobuhiko  
KURIHARA, Tatsuya  
YAMAICHI, Kozo  
YAMAGUCHI, Nozomi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/091,028A  
FILING DATE: 14-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: REA, TERESA STANEK



REGISTRATION NUMBER: 30,427  
REFERENCE/DOCKET NUMBER: 001560-204  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-6620  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-091-028A-34

Query Match  
Best Local Similarity 100.0%; Score 52; DB 4; Length 380;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SETINCHFK 9  
DB 188 SETINCHFK 196

RESULT 4  
US-09-140-719-34  
Sequence 34, Application US/09140719  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSUROUKA, Nobuo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, Nobuhiko  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMACHU, Kozo  
APPLICANT: YAMAGUCHI, Nozomi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 34:

SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-140-719-34

Query Match  
Best Local Similarity 100.0%; Score 52; DB 15; Length 380;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SETINCHFK 9  
DB 188 SETINCHFK 196

RESULT 5  
US-09-508-997A-2  
Sequence 2, Application US/09508997A  
GENERAL INFORMATION:  
APPLICANT: MIYATA, Toshio  
TITLE OF INVENTION: Megsin Protein  
FILE REFERENCE: SHIM-004  
CURRENT APPLICATION NUMBER: US/09/508,997A  
CURRENT FILING DATE: 2000-06-27  
PRIOR APPLICATION NUMBER: JP 09-275302  
PRIOR FILING DATE: 1997-09-22  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 2  
LENGTH: 380  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-508-997A-2

Query Match  
Best Local Similarity 100.0%; Score 52; DB 19; Length 380;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SETINCHFK 9  
DB 188 SETINCHFK 196

RESULT 6  
US-09-791-537-281  
Sequence 281, Application US/09791537  
GENERAL INFORMATION:  
APPLICANT: Bionomix, Inc.  
APPLICANT: Debe, Derek  
APPLICANT: Danzer, Joseph  
TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME  
FILE REFERENCE: 261/210  
CURRENT APPLICATION NUMBER: US/09/791,537  
CURRENT FILING DATE: 2001-02-22  
NUMBER OF SEQ ID NOS: 153055  
SOFTWARE: Patent In version 3.0  
SEQ ID NO 281  
LENGTH: 380  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-791-537-281

Query Match  
Best Local Similarity 100.0%; Score 52; DB 21; Length 380;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SETINCHFK 9  
DB 188 SETINCHFK 196

```
RESULT 7
US-09-936-883A-2
; Sequence 2, Application US/09936883A
; GENERAL INFORMATION:
; APPLICANT: MIYATA, Toshio
; TITLE OF INVENTION: A Method for Detecting Megsin Protein and Use
; FILE REFERENCE: F2-10IDPICT
; CURRENT APPLICATION NUMBER: US/09/936, 883A
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: JP 1999-75305
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: JP 1999-306623
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-936-883A-2

Query Match      100.0%; Score 52; DB 23; Length 380;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 SETINCHK 9
Db      188 SETINCHK 196

RESULT 8
US-10-091-442-34
; Sequence 34, Application US/10091442
; GENERAL INFORMATION:
; APPLICANT: TSUJIMOTO, Masafumi
; TWASA, Fuyuki
; TSURUOKA, Nobuo
; NAKAZATO, Hiroshi
; MIURA, Kenju
; ISHIDA, Nobuhiko
; KURIHARA, Tatsuya
; YAMAGUCHI, Kozo
; YAMAGUCHI, Nozomi
; TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/091,442
; FILING DATE: 07-Mar-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/140,719
; FILING DATE: 08-AUG-1998
; APPLICATION NUMBER: US 08/474,661
; FILING DATE: 07-JUN-1995
; APPLICATION NUMBER: US 08/091,028
; FILING DATE: 14-JUL-1993
; APPLICATION NUMBER: JP 4-212305
; FILING DATE: 17-JUL-1992
; APPLICATION NUMBER: JP 6-067339
; FILING DATE: 04-MAR-1993
```

```
ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm R.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 001560-247
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 380 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 34:
US-10-091-442-34

Query Match      100.0%; Score 52; DB 24; Length 380;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 SETINCHK 9
Db      188 SETINCHK 196

RESULT 9
US-10-110-030-2
; Sequence 2, Application US/10110030
; GENERAL INFORMATION:
; APPLICANT: Miyata, Toshio
; TITLE OF INVENTION: ANIMAL MODEL FOR MESENTERIAL PROLIFERATIVE
; FILE REFERENCE: SHIM015
; CURRENT APPLICATION NUMBER: US/10/110, 030
; PRIOR FILING DATE: 2002-04-03
; PRIOR APPLICATION NUMBER: 11/285736
; PRIOR FILING DATE: 1999-10-06
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-110-030-2

Query Match      100.0%; Score 52; DB 25; Length 380;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 SETINCHK 9
Db      188 SETINCHK 196

RESULT 10
US-09-724-059-151732
; Sequence 151732, Application US/09724059
; GENERAL INFORMATION:
; APPLICANT: Choo, Yen
; APPLICANT: Klug, Aaron
; APPLICANT: Isalan, Mark
; TITLE OF INVENTION: Nucleic Acid Binding Proteins
; FILE REFERENCE: P2500USM
; CURRENT APPLICATION NUMBER: US/09/724,059
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: GB9710809.6
; PRIOR FILING DATE: 1997-05-23
; NUMBER OF SEQ ID NOS: 1407122
; SOFTWARE: MacAlignster
; SEQ ID NO 151732
; LENGTH: 23
; TYPE: PRT
; ORGANISM: Artificial sequence
```

FEATURE:  
OTHER INFORMATION: Polypeptide sequence  
US-09-724-059-151732

Query Match 75.0%: Score 39; DB 21; Length 23;  
Best Local Similarity 66.7%: Pred. No. 16;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 SETINCHK 9  
DB 13 SETLSCHK 21

RESULT 11  
US-09-724-059-154252  
Sequence 154252, Application US/09724059  
GENERAL INFORMATION:  
APPLICANT: Choo, Yen  
APPLICANT: Klug, Aaron  
TITLE OF INVENTION: Nucleic Acid Binding Proteins  
FILE REFERENCE: P2500USM  
CURRENT FILING DATE: 2000-11-28  
PRIOR APPLICATION NUMBER: GB9710809.6  
PRIOR FILING DATE: 1997-05-23  
NUMBER OF SEQ ID NOS: 1407122  
SOFTWARE: Macallister  
SEQ ID NO 154252  
LENGTH: 23  
TYPE: PRT  
ORGANISM: Artificial sequence  
FEATURE:  
OTHER INFORMATION: Polypeptide sequence  
US-09-724-059-154252

Query Match 75.0%: Score 39; DB 21; Length 23;  
Best Local Similarity 66.7%: Pred. No. 16;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 SETINCHK 9  
DB 13 SETLSCHK 21

RESULT 12  
US-09-724-059-156652  
Sequence 156652, Application US/09724059  
GENERAL INFORMATION:  
APPLICANT: Choo, Yen  
APPLICANT: Klug, Aaron  
APPLICANT: Isalan, Mark  
TITLE OF INVENTION: Nucleic Acid Binding Proteins  
FILE REFERENCE: P2500USM  
CURRENT APPLICATION NUMBER: US/09/724,059  
CURRENT FILING DATE: 2000-11-28  
PRIOR APPLICATION NUMBER: GB9710809.6  
PRIOR FILING DATE: 1997-05-23  
NUMBER OF SEQ ID NOS: 1407122  
SOFTWARE: Macallister  
SEQ ID NO 156652  
LENGTH: 23  
TYPE: PRT  
ORGANISM: Artificial sequence  
FEATURE:  
OTHER INFORMATION: Polypeptide sequence  
US-09-724-059-156652

Query Match 75.0%: Score 39; DB 21; Length 23;  
Best Local Similarity 66.7%: Pred. No. 16;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 SETINCHK 9

DB 13 SETLSCHK 21

RESULT 13  
US-09-724-059-159172  
Sequence 159172, Application US/09724059  
GENERAL INFORMATION:  
APPLICANT: Choo, Yen  
APPLICANT: Klug, Aaron  
APPLICANT: Isalan, Mark  
TITLE OF INVENTION: Nucleic Acid Binding Proteins  
FILE REFERENCE: P2500USM  
CURRENT APPLICATION NUMBER: US/09/724,059  
CURRENT FILING DATE: 2000-11-28  
PRIOR APPLICATION NUMBER: GB9710809.6  
PRIOR FILING DATE: 1997-05-23  
NUMBER OF SEQ ID NOS: 1407122  
SOFTWARE: Macallister  
SEQ ID NO 159172  
LENGTH: 23  
TYPE: PRT  
ORGANISM: Artificial sequence  
FEATURE:  
OTHER INFORMATION: Polypeptide sequence  
US-09-724-059-159172

Query Match 75.0%: Score 39; DB 21; Length 23;  
Best Local Similarity 66.7%: Pred. No. 16;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 SETINCHK 9  
DB 13 SETLSCHK 21

RESULT 14  
US-09-724-059-161692  
Sequence 161692, Application US/09724059  
GENERAL INFORMATION:  
APPLICANT: Choo, Yen  
APPLICANT: Klug, Aaron  
APPLICANT: Isalan, Mark  
TITLE OF INVENTION: Nucleic Acid Binding Proteins  
FILE REFERENCE: P2500USM  
CURRENT APPLICATION NUMBER: US/09/724,059  
CURRENT FILING DATE: 2000-11-28  
PRIOR APPLICATION NUMBER: GB9710809.6  
PRIOR FILING DATE: 1997-05-23  
NUMBER OF SEQ ID NOS: 1407122  
SOFTWARE: Macallister  
SEQ ID NO 161692  
LENGTH: 23  
TYPE: PRT  
ORGANISM: Artificial sequence  
FEATURE:  
OTHER INFORMATION: Polypeptide sequence  
US-09-724-059-161692

Query Match 75.0%: Score 39; DB 21; Length 23;  
Best Local Similarity 66.7%: Pred. No. 16;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 SETINCHK 9  
DB 13 SETLSCHK 21

RESULT 15  
US-09-724-059-163972  
Sequence 163972, Application US/09724059  
GENERAL INFORMATION:  
APPLICANT: Choo, Yen

APPLICANT: Klug, Aaron  
APPLICANT: Isalan, Mark  
TITLE OF INVENTION: Nucleic Acid Binding Proteins  
FILE REFERENCE: P2500USM  
CURRENT APPLICATION NUMBER: US/09/724,059  
CURRENT FILING DATE: 2000-11-28  
PRIOR APPLICATION NUMBER: GB9710809.6  
PRIOR FILING DATE: 1997-05-23  
NUMBER OF SEQ ID NOS: 1407122  
SOFTWARE: MacAllister  
SEQ ID NO: 163972  
LENGTH: 23  
TYPE: PRT  
ORGANISM: Artificial sequence  
FEATURE:  
OTHER INFORMATION: Polypeptide sequence  
US-09-724-059-163972

Query Match 75.0%; Score 39; DB 21; Length 23;  
Best Local Similarity 66.7%; Pred. No. 16;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SETINCHPK 9  
|||::|||  
DB 13 SETLSCHTK 21

Search completed: July 11, 2003, 12:27:26  
Job time : 31.3285 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:55:27 ; Search time 8.45738 Seconds

(without alignments)  
280.699 Million cell updates/sec

Title: US-10-091-442-1

Perfect score: 52

Sequence: 1 SETINCHFX 9

Scoring table: BLOSUM62

Searched: 1229476 seqs, 263775074 residues

Total number of hits satisfying chosen parameters: 1229476

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Pending\_Patents\_AA\_New:\*  
1: /cgn2\_6/ptodata/1/paa/PCF\_NEW\_COMB.pep4:\*  
2: /cgn2\_6/ptodata/1/paa/PCF\_NEW\_COMB.pep4:\*  
3: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pep4:\*  
4: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep4:\*  
5: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep4:\*  
6: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep4:\*  
7: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep4:\*  
8: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep4:\*  
9: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep4:\*  
10: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep4:\*  
11: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep4:\*  
12: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep4:\*  
13: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep4:\*  
14: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	52	100.0	380	10	US-09-936-883C-2
2	52	100.0	380	12	US-10-126-052A-263
3	52	100.0	380	14	US-60-443-566-3006
4	52	100.0	380	14	US-60-455-444-4692
5	52	100.0	380	14	US-60-465-241-4692
6	40	76.9	53	12	US-10-424-599-202788
7	39	75.0	380	10	US-09-936-883C-19
8	37	71.2	56	12	US-10-424-599-158716
9	37	71.2	275	12	US-10-425-114-47479
10	37	71.2	320	12	US-10-219-051B-13364
11	37	71.2	638	12	US-10-425-114-71838
12	37	71.2	1030	2	PCT-US01-29288A-140
13	37	71.2	1030	12	US-10-149-310-140
14	36	69.2	86	12	US-10-424-599-223211
15	36	69.2	239	12	US-10-425-114-58872
16	36	69.2	421	12	US-10-425-114-57566
17	36	69.2	863	2	PCT-US01-29288A-264
18	36	69.2	863	12	US-10-149-310-264
19	36	69.2	1007	12	US-10-425-114-62667

20	36	69.2	1215	12	US-10-437-963-129618	Sequence 129618, A
21	35	67.3	309	12	US-10-425-114-51109	Sequence 51109, A
22	35	67.3	522	12	US-10-282-122A-73834	Sequence 73834, A
23	35	67.3	858	12	US-10-369-294-18	Sequence 18, App1
24	35	67.3	864	12	US-10-437-963-153309	Sequence 153309, A
25	35	67.3	1173	12	US-10-369-493-5025	Sequence 5025, App
26	34	65.4	56	12	US-10-424-599-192457	Sequence 192457, A
27	34	65.4	75	12	US-10-425-114-50539	Sequence 50539, A
28	34	65.4	117	12	US-10-425-114-40164	Sequence 40164, A
29	34	65.4	120	12	US-10-416-991-111	Sequence 111, App
30	34	65.4	123	12	US-10-416-991-110	Sequence 110, App
31	34	65.4	127	12	US-10-417-884-5938	Sequence 5938, App
32	34	65.4	201	12	US-10-125-923A-608	Sequence 608, App
33	34	65.4	201	12	US-10-223-089-344	Sequence 344, App
34	34	65.4	201	12	US-10-205-892-608	Sequence 608, App
35	34	65.4	201	12	US-10-174-575-608	Sequence 608, App
36	34	65.4	201	12	US-10-187-755-608	Sequence 608, App
37	34	65.4	201	12	US-10-187-749-608	Sequence 608, App
38	34	65.4	201	12	US-10-199-672-608	Sequence 608, App
39	34	65.4	201	12	US-10-184-486-608	Sequence 608, App
40	34	65.4	201	12	US-10-305-654-344	Sequence 344, App1
41	34	65.4	215	12	US-10-360-186-54	Sequence 54, App1
42	34	65.4	292	12	US-10-425-114-39749	Sequence 39749, A
43	34	65.4	292	12	US-10-425-114-40007	Sequence 40007, A
44	34	65.4	292	12	US-10-424-599-183093	Sequence 183093, A
45	34	65.4	292	12	US-10-424-599-183093	Sequence 183093, A

#### ALIGNMENTS

RESULT 1  
US-09-936-883C-2  
; Sequence 2, Application US/09936883C  
; GENERAL INFORMATION:  
; APPLICANT: MIYATA, Toshio  
; TITLE OF INVENTION: A Method for Detecting Megsln Protein and Use  
; TITLE OF INVENTION: Thereof  
; FILE REFERENCE: F2-10DIPICF  
; CURRENT APPLICATION NUMBER: US/09/936,883C  
; PRIOR FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: JP 1999-75305  
; PRIOR FILING DATE: 1999-03-19  
; PRIOR APPLICATION NUMBER: JP 1999-306623  
; PRIOR FILING DATE: 1999-10-28  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: Patentln Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 380  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-936-883C-2

Query Match  
Best Local Similarity 100.0%; Score 52; DB 10; Length 380;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SETINCHFX 9  
DB 188 SETINCHFX 196

RESULT 2  
US-10-126-052A-263  
; Sequence 263, Application US/10126052A  
; GENERAL INFORMATION:  
; APPLICANT: Aziz, Natasha  
; APPLICANT: Murray, Richard  
; APPLICANT: Eos Biotechnology, Inc.  
; TITLE OF INVENTION: Methods of Diagnosis of Lung Cancer. Compositions and  
; TITLE OF INVENTION: Methods of Screening for Modulators of Lung Cancer  
; FILE REFERENCE: 018501-001530US  
; CURRENT APPLICATION NUMBER: US/10/126,052A

```
;; CURRENT FILING DATE: 2002-04-18
;; PRIOR APPLICATION NUMBER: US 60/284,770
;; PRIOR FILING DATE: 2001-04-18
;; PRIOR APPLICATION NUMBER: US 60/290,492
;; PRIOR FILING DATE: 2001-05-10
;; PRIOR APPLICATION NUMBER: US 60/339,245
;; PRIOR FILING DATE: 2001-11-09
;; PRIOR APPLICATION NUMBER: US 60/350,666
;; PRIOR FILING DATE: 2001-11-13
;; PRIOR APPLICATION NUMBER: US 60/334,370
;; PRIOR FILING DATE: 2001-11-29
;; PRIOR APPLICATION NUMBER: US 60/372,246
;; PRIOR FILING DATE: 2002-04-12
;; NUMBER OF SEQ ID NOS: 691
;; SOFTWARE: Patentl Ver. 2.1
;; SEQ ID NO 263
;; LENGTH: 380
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-126-052A-263
```

```
Query Match
Best Local Similarity 100.0%; Score 52; DB 12; Length 380;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 SETINCHK 9
Db 188 SETINCHK 196
```

```
RESULT 3
US-60-443-566-3006
;; Sequence 3006, Application US/60443566
;; GENERAL INFORMATION:
;; APPLICANT: CARGILL, Michele
;; APPLICANT: BEGOVICH, Ann
;; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;; FILE REFERENCE: C1001447
;; CURRENT APPLICATION NUMBER: US/60/443,566
;; CURRENT FILING DATE: 2003-01-30
;; NUMBER OF SEQ ID NOS: 25102
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 3006
;; LENGTH: 380
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-60-443-566-3006
```

```
Query Match
Best Local Similarity 100.0%; Score 52; DB 14; Length 380;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 SETINCHK 9
Db 188 SETINCHK 196
```

```
RESULT 4
US-60-455-444-4692
;; Sequence 4692, Application US/60455444
;; GENERAL INFORMATION:
;; APPLICANT: CARGILL, Michele
;; APPLICANT: BEGOVICH, Ann
;; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;; FILE REFERENCE: C1001455
;; CURRENT APPLICATION NUMBER: US/60/455,444
;; CURRENT FILING DATE: 2003-03-18
;; NUMBER OF SEQ ID NOS: 50986
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 4692
;; LENGTH: 380
```

```
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-60-455-444-4692
```

```
Query Match
Best Local Similarity 100.0%; Score 52; DB 14; Length 380;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 SETINCHK 9
Db 188 SETINCHK 196
```

```
RESULT 5
US-60-465-241-4692
;; Sequence 4692, Application US/60465241
;; GENERAL INFORMATION:
;; APPLICANT: CARGILL, Michele
;; APPLICANT: BEGOVICH, Ann
;; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;; FILE REFERENCE: C1001468
;; CURRENT APPLICATION NUMBER: US/60/465,241
;; CURRENT FILING DATE: 2003-04-23
;; NUMBER OF SEQ ID NOS: 258418
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 4692
;; LENGTH: 380
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-60-465-241-4692
```

```
Query Match
Best Local Similarity 100.0%; Score 52; DB 14; Length 380;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 SETINCHK 9
Db 188 SETINCHK 196
```

```
RESULT 6
US-10-424-599-202788
;; Sequence 202788, Application US/10424599
;; GENERAL INFORMATION:
;; APPLICANT: La Rosa Thomas J
;; APPLICANT: Kovalic David K
;; APPLICANT: Zhou Yihua
;; TITLE OF INVENTION: Soy, Nucleic Acid Molecules and Other Molecules Associated With
;; FILE REFERENCE: 38-21(53223)B
;; CURRENT APPLICATION NUMBER: US/10/424,599
;; CURRENT FILING DATE: 2003-04-28
;; NUMBER OF SEQ ID NOS: 285684
;; SEQ ID NO 202788
;; LENGTH: 53
;; TYPE: PRT
;; ORGANISM: Glycine max
;; FEATURE:
;; NAME/KEY: unsure
;; LOCATION: (1)..(53)
;; OTHER INFORMATION: unsure at all Xaa locations
;; FEATURE:
;; OTHER INFORMATION: Clone ID: PAT_MKT3847_25142C.1.pcp
US-10-424-599-202788
```

```
Query Match
Best Local Similarity 76.9%; Score 40; DB 12; Length 53;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 2 ETINCHK 9
: |||||
```



APPLICANT: Tabaska, Jack E  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(5313)B  
CURRENT APPLICATION NUMBER: US/10/425,114  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 73128  
SEQ ID NO 71838  
LENGTH: 638  
TYPE: PRT  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: L1B3245-495-B7\_FLI.pep  
US-10-425-114-71838

Query Match  
Best Local Similarity 71.2%; Score 37; DB 12; Length 638;  
Best Local Similarity 66.7%; Pred. No. 1.2e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 SETINCHK 9  
Db 31 SETINCHME 39

RESULT 12  
PCT-US01-29288A-140  
Sequence 140, Application PC/TUS0129288A  
GENERAL INFORMATION:  
APPLICANT: Microbia, Inc.  
TITLE OF INVENTION: Modulation of Secondary Metabolite Production by  
TITLE OF INVENTION: Zinc Binuclear Cluster Proteins  
FILE REFERENCE: 14184-019W01  
CURRENT APPLICATION NUMBER: PCT/US01/29288A  
CURRENT FILING DATE: 2001-09-19  
PRIOR APPLICATION NUMBER: US 60/233,564  
PRIOR FILING DATE: 2000-09-19  
NUMBER OF SEQ ID NOS: 308  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 140  
LENGTH: 1030  
TYPE: PRT  
ORGANISM: Saccharomyces cerevisiae  
PCT-US01-29288A-140

Query Match  
Best Local Similarity 71.2%; Score 37; DB 2; Length 1030;  
Best Local Similarity 55.6%; Pred. No. 1.8e+02;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 SETINCHK 9  
Db 689 SRVLNCHFR 697

RESULT 13  
US-10-149-310-140  
Sequence 140, Application US/10149310  
GENERAL INFORMATION:  
APPLICANT: Holtzman, Douglas  
APPLICANT: Madden, Kevin T.  
APPLICANT: Maxon, Mary  
APPLICANT: Sherman, Amir  
TITLE OF INVENTION: Modulation of Secondary Metabolite Production by  
TITLE OF INVENTION: Zinc Binuclear Cluster Proteins  
FILE REFERENCE: 14184-019W01  
CURRENT APPLICATION NUMBER: US/10/149,310  
CURRENT FILING DATE: 2003-02-19  
PRIOR APPLICATION NUMBER: PCT/US01/29288  
PRIOR FILING DATE: 2001-09-19  
PRIOR APPLICATION NUMBER: US 60/233,564  
PRIOR FILING DATE: 2000-09-19  
NUMBER OF SEQ ID NOS: 308  
SOFTWARE: PatentIn version 3.1

SEQ ID NO 140  
LENGTH: 1030  
TYPE: PRT  
ORGANISM: Saccharomyces cerevisiae  
US-10-149-310-140

Query Match  
Best Local Similarity 71.2%; Score 37; DB 12; Length 1030;  
Best Local Similarity 55.6%; Pred. No. 1.8e+02;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 SETINCHK 9  
Db 689 SRVLNCHFR 697

RESULT 14  
US-10-424-599-223211  
Sequence 223211, Application US/10424599  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 223211  
LENGTH: 86  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_43590C.1.pep  
US-10-424-599-223211

Query Match  
Best Local Similarity 69.2%; Score 36; DB 12; Length 86;  
Best Local Similarity 83.3%; Pred. No. 41;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 TINCH 8  
Db 56 TLNCHF 61

RESULT 15  
US-10-425-114-58872  
Sequence 58872, Application US/10425114  
GENERAL INFORMATION:  
APPLICANT: Liu, Jingdong  
APPLICANT: Zhou Yihua  
APPLICANT: Kovalic, David K.  
APPLICANT: Screen, Steven E  
APPLICANT: Tabaska, Jack E  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(5313)B  
CURRENT APPLICATION NUMBER: US/10/425,114  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 73128  
SEQ ID NO 58872  
LENGTH: 239  
TYPE: PRT  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: 700805614\_FLI.pep  
US-10-425-114-58872

Query Match  
Best Local Similarity 69.2%; Score 36; DB 12; Length 239;  
Best Local Similarity 62.5%; Pred. No. 89;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;



Fri Jul 11 15:10:33 2003

us-10-091-442-1.rpn

Page 5

Oy	2	ETINCHFR	9
	11	:	111
Db	23	ETVMCHFR	30

Search completed: July 11, 2003, 12:35:04  
Job time : 9.55738 secs

---

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:54:32 ; Search time 22.0333 Seconds  
(without alignments)  
204.833 Million cell updates/sec

Title: US-10-091-442-2  
Perfect score: 40  
Sequence: 1 WQSAFTK 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 64473110 residues  
Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_AA\_Main:\*  
1: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/paa/US087\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/paa/US087\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/paa/US088\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/paa/US088\_COMB.pep.\*  
7: /cgn2\_6/ptodata/1/paa/US088\_COMB.pep.\*  
8: /cgn2\_6/ptodata/1/paa/US088\_COMB.pep.\*  
9: /cgn2\_6/ptodata/1/paa/US088\_COMB.pep.\*  
10: /cgn2\_6/ptodata/1/paa/US088\_COMB.pep.\*  
11: /cgn2\_6/ptodata/1/paa/US088\_COMB.pep.\*  
12: /cgn2\_6/ptodata/1/paa/US088\_COMB.pep.\*  
13: /cgn2\_6/ptodata/1/paa/US089\_COMB.pep.\*  
14: /cgn2\_6/ptodata/1/paa/US090\_COMB.pep.\*  
15: /cgn2\_6/ptodata/1/paa/US091\_COMB.pep.\*  
16: /cgn2\_6/ptodata/1/paa/US092\_COMB.pep.\*  
17: /cgn2\_6/ptodata/1/paa/US093\_COMB.pep.\*  
18: /cgn2\_6/ptodata/1/paa/US094\_COMB.pep.\*  
19: /cgn2\_6/ptodata/1/paa/US095\_COMB.pep.\*  
20: /cgn2\_6/ptodata/1/paa/US097\_COMB.pep.\*  
21: /cgn2\_6/ptodata/1/paa/US097\_COMB.pep.\*  
22: /cgn2\_6/ptodata/1/paa/US098\_COMB.pep.\*  
23: /cgn2\_6/ptodata/1/paa/US099\_COMB.pep.\*  
24: /cgn2\_6/ptodata/1/paa/US100\_COMB.pep.\*  
25: /cgn2\_6/ptodata/1/paa/US101\_COMB.pep.\*  
26: /cgn2\_6/ptodata/1/paa/US102\_COMB.pep.\*  
27: /cgn2\_6/ptodata/1/paa/US60\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40	100.0	7	US-09-140-719-2	Sequence 2, Appli
2	40	100.0	7	US-10-091-442-2	Sequence 2, Appli
3	40	100.0	380	US-08-091-028A-34	Sequence 34, Appli
4	40	100.0	380	US-09-140-719-34	Sequence 34, Appli
5	40	100.0	380	US-09-508-997A-21	Sequence 2, Appli
6	40	100.0	380	US-09-791-537-281	Sequence 281, App

7	40	100.0	380	US-09-936-883A-2	Sequence 2, Appli
8	40	100.0	380	US-10-091-442-34	Sequence 34, Appli
9	40	100.0	380	US-10-110-030-2	Sequence 2, Appli
10	36	90.0	306	PCT-US01-08631-33136	Sequence 33136, A
11	36	90.0	323	PCT-US01-08631-41856	Sequence 41856, A
12	36	90.0	368	US-09-508-997A-6	Sequence 6, Appli
13	36	90.0	368	US-09-936-883A-21	Sequence 21, Appli
14	36	90.0	368	US-10-110-030-6	Sequence 6, Appli
15	36	90.0	380	US-09-508-997A-4	Sequence 4, Appli
16	36	90.0	380	US-09-936-883A-19	Sequence 19, Appli
17	36	90.0	380	US-10-110-030-4	Sequence 4, Appli
18	35	87.5	85	US-60-160-202-2229	Sequence 2229, Ap
19	35	87.5	85	US-60-169-842-2638	Sequence 2638, Ap
20	35	87.5	95	US-60-164-763-713	Sequence 713, App
21	35	87.5	95	US-60-169-842-4466	Sequence 4466, Ap
22	35	87.5	139	PCT-US01-14827-11233	Sequence 11233, A
23	35	87.5	586	US-10-104-047-2592	Sequence 2592, Ap
24	34	85.0	54	US-60-128-476-3191	Sequence 3191, Ap
25	32	80.0	95	US-60-160-202-3165	Sequence 3165, Ap
26	32	80.0	127	US-60-169-842-4731	Sequence 4731, Ap
27	32	80.0	219	PCT-US01-08631-60256	Sequence 60256, A
28	32	80.0	331	PCT-US01-08631-48701	Sequence 48701, A
29	32	80.0	369	US-60-360-039-12166	Sequence 12166, A
30	32	80.0	404	US-10-108-260A-4601	Sequence 4601, Ap
31	32	80.0	736	US-09-791-537-70938	Sequence 70938, A
32	32	80.0	931	US-09-830-446-43	Sequence 43, Appli
33	32	80.0	940	US-09-198-452A-500	Sequence 500, App
34	32	80.0	940	US-09-438-185-468	Sequence 468, App
35	32	80.0	940	US-09-438-185A-468	Sequence 468, App
36	32	80.0	944	PCT-US01-08656-9504	Sequence 9504, Ap
37	32	80.0	944	US-60-389-987-2257	Sequence 2257, Ap
38	32	80.0	944	US-60-412-418-2257	Sequence 2257, Ap
39	32	80.0	959	US-10-109-268A-12	Sequence 12, Appli
40	32	80.0	1072	US-09-614-150-13284	Sequence 13284, A
41	32	80.0	1072	US-60-191-637-13325	Sequence 13325, A
42	32	80.0	1072	US-60-191-681-10493	Sequence 10493, A
43	32	80.0	1074	US-60-167-217-13330	Sequence 13330, A
44	32	80.0	1074	US-60-173-464-10850	Sequence 10850, A
45	31	77.5	46	PCT-US00-06012-112	Sequence 112, App

## ALIGNMENTS

RESULT 1  
US-09-140-719-2  
; Sequence 2, Application US/09140719  
; GENERAL INFORMATION:  
; APPLICANT: TSUJIMOTO, Masatumi  
; APPLICANT: IWASA, Fuyuki  
; APPLICANT: TSURUOKA, Nobuo  
; APPLICANT: NAKAZATO, Hiroshi  
; APPLICANT: MURA, Kenju  
; APPLICANT: ISHIDA, Nobuhiko  
; APPLICANT: KURIHARA, Tatsuya  
; APPLICANT: YAMACHI, Kozo  
; APPLICANT: YAMAGUCHI, Nozomi  
; TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
; NUMBER OF SEQUENCES: 34  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: P. O. Box 1404  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/140,719

FILING DATE: 08-AUG-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-140-719-2

Query Match 100.0%; Score 40; DB 15; Length 7;  
Best Local Similarity 100.0%; Pred. No. 4.2e+06;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WOSAFTK 7  
Db 1 WOSAFTK 7

RESULT 2  
US-10-091-442-2  
Sequence 2, Application US/10091442  
GENERAL INFORMATION:  
APPLICANT: TSUTIMOTO, Masafumi  
IMASA, Fuyuki  
TSURUOKA, Nobuo  
NAKAZATO, Hiroshi  
MIURA, Kenju  
ISHIDA, Nobuhito  
KURIHARA, Tatsuya  
YAMAICHI, Kozo  
YAMAGUCHI, Nozomi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/091,442  
FILING DATE: 07-Mar-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998

APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-10-091-442-2

Query Match 100.0%; Score 40; DB 24; Length 7;  
Best Local Similarity 100.0%; Pred. No. 4.2e+06;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WOSAFTK 7  
Db 1 WOSAFTK 7

RESULT 3  
US-08-091-028A-34  
Sequence 34, Application US/08091028A  
GENERAL INFORMATION:  
APPLICANT: TSUTIMOTO, Masafumi  
IMASA, Fuyuki  
TSURUOKA, Nobuo  
NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, Nobuhito  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, Nozomi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/091,028A  
FILING DATE: 14-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: REA, TERESA STANER

REGISTRATION NUMBER: 30,427  
REFERENCE/DOCKET NUMBER: 001560-204  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-6620  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-091-028A-34

Query Match 100.0%; Score 40; DB 4; Length 380;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WOSAFTK 7  
Db 181 WOSAFTK 187

RESULT 4  
US-09-140-719-34  
Sequence 34, Application US/09140719  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSUROUOKA, Nobuo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenji  
APPLICANT: ISHIDA, Nobuhiko  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAGUCHI, Nozomi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 34:

SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-140-719-34

Query Match 100.0%; Score 40; DB 15; Length 380;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WOSAFTK 7  
Db 181 WOSAFTK 187

RESULT 5  
US-09-508-997A-2  
Sequence 2, Application US/09508997A  
GENERAL INFORMATION:  
APPLICANT: MIYATA, Toshio  
TITLE OF INVENTION: Megsin Protein  
FILE REFERENCE: SHIM-004  
CURRENT APPLICATION NUMBER: US/09/508,997A  
CURRENT FILING DATE: 2000-06-27  
PRIOR APPLICATION NUMBER: JP 09-275302  
PRIOR FILING DATE: 1997-09-22  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 380  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-508-997A-2

Query Match 100.0%; Score 40; DB 19; Length 380;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WOSAFTK 7  
Db 181 WOSAFTK 187

RESULT 6  
US-09-791-537-281  
Sequence 281, Application US/09791537  
GENERAL INFORMATION:  
APPLICANT: Bionomix, Inc.  
APPLICANT: Debe, Derek  
TITLE OF INVENTION: Danzer, Joseph  
TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME  
FILE REFERENCE: 261/210  
CURRENT APPLICATION NUMBER: US/09/791,537  
CURRENT FILING DATE: 2001-02-22  
NUMBER OF SEQ ID NOS: 153055  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 281  
LENGTH: 380  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-791-537-281

Query Match 100.0%; Score 40; DB 21; Length 380;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WOSAFTK 7  
Db 181 WOSAFTK 187

RESULT 7  
US-09-936-883A-2  
Sequence 2, Application US/09936883A  
GENERAL INFORMATION:  
APPLICANT: MIYATA, Toshio  
TITLE OF INVENTION: A Method for Detecting Megsin Protein and Use  
TITLE OF INVENTION: " thereof  
FILE REFERENCE: F2-101DIPCT  
CURRENT APPLICATION NUMBER: US/09/936, 883A  
CURRENT FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: JP 1999-75305  
PRIOR FILING DATE: 1999-03-19  
PRIOR APPLICATION NUMBER: JP 1999-306623  
PRIOR FILING DATE: 1999-10-28  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 380  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-936-883A-2

Query Match  
Best Local Similarity 100.0%; Score 40; DB 23; Length 380;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WSAFTK 7  
181 WSAFTK 187

RESULT 8  
US-10-091-442-34  
Sequence 34, Application US/10091442  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
IMASA, Fuyuki  
TSUKUOKA, Nobuo  
NAKAZATO, Hiroshi  
MIURA, Kenju  
ISHIDA, Nobuhiko  
KURIHARA, Tatsuya  
YAMAICHI, Kozo  
YAMAGUCHI, Nozomi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 2213-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/091,442  
FILING DATE: 07-Mar-2002  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993

ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
MOLECULE TYPE: protein  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 34:  
US-10-091-442-34

Query Match  
Best Local Similarity 100.0%; Score 40; DB 24; Length 380;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WSAFTK 7  
181 WSAFTK 187

RESULT 9  
US-10-110-030-2  
Sequence 2, Application US/10110030  
GENERAL INFORMATION:  
APPLICANT: Miyata, Toshio  
TITLE OF INVENTION: ANIMAL MODEL FOR MESENTERIAL PROLIFERATIVE  
FILE REFERENCE: SHIM015  
CURRENT APPLICATION NUMBER: US/10/110,030  
CURRENT FILING DATE: 2002-04-03  
PRIOR APPLICATION NUMBER: 11/285736  
PRIOR FILING DATE: 1999-10-06  
NUMBER OF SEQ ID NOS: 15  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 380  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-110-030-2

Query Match  
Best Local Similarity 100.0%; Score 40; DB 25; Length 380;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WSAFTK 7  
181 WSAFTK 187

RESULT 10  
PCT-US01-08631-33136  
Sequence 33136, Application PC/TUS0108631  
GENERAL INFORMATION:  
APPLICANT: Hyseq, Inc  
TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES  
FILE REFERENCE: 21272-049  
CURRENT APPLICATION NUMBER: PCT/US01/08631  
CURRENT FILING DATE: 2001-03-30  
PRIOR APPLICATION NUMBER: 09/540,217  
PRIOR FILING DATE: 2000-03-31  
PRIOR APPLICATION NUMBER: 09/649,167  
PRIOR FILING DATE: 2000-08-23  
NUMBER OF SEQ ID NOS: 60736  
SOFTWARE: Custom  
SEQ ID NO 33136  
LENGTH: 306  
TYPE: PRT  
ORGANISM: Homo sapiens

FEATURE:  
NAME/KEY: misc.feature  
LOCATION: (1)-(306)  
OTHER INFORMATION: Xaa - x or \* as defined in Table 2  
PCT-US01-08631-33136

Query Match 90.0%; Score 36; DB 1; Length 306;  
Best Local Similarity 85.7%; Pred. No. 1.9e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 WOSAFTK 7  
111111  
Db 108 WOSVFTK 114

RESULT 11  
PCT-US01-08631-41856  
Sequence 41856, Application PC/TUS0108631  
GENERAL INFORMATION:  
APPLICANT: Hyseq, Inc  
TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES  
FILE REFERENCE: 21272-049  
CURRENT APPLICATION NUMBER: PCT/US01/08631  
CURRENT FILING DATE: 2001-03-30  
PRIOR APPLICATION NUMBER: 09/540,217  
PRIOR FILING DATE: 2000-03-31  
PRIOR APPLICATION NUMBER: 09/649,167  
PRIOR FILING DATE: 2000-08-23  
NUMBER OF SEQ ID NOS: 60736  
SOFTWARE: Custom  
SEQ ID NO 41856  
LENGTH: 323  
TYPE: PRT  
ORGANISM: Homo sapiens  
PCT-US01-08631-41856

Query Match 90.0%; Score 36; DB 1; Length 323;  
Best Local Similarity 85.7%; Pred. No. 1.9e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 WOSAFTK 7  
111111  
Db 108 WOSVFTK 114

RESULT 12  
US-09-508-997A-6  
Sequence 6, Application US/09508997A  
GENERAL INFORMATION:  
APPLICANT: MIYATA, Toshio  
TITLE OF INVENTION: Megsin Protein  
FILE REFERENCE: SHIM-004  
CURRENT APPLICATION NUMBER: US/09/508,997A  
CURRENT FILING DATE: 2000-06-27  
PRIOR APPLICATION NUMBER: JP 09-275302  
PRIOR FILING DATE: 1997-09-22  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: Patentln Ver. 2.0  
SEQ ID NO 6  
LENGTH: 368  
TYPE: PRT  
ORGANISM: Mus musculus  
US-09-508-997A-6

Query Match 90.0%; Score 36; DB 19; Length 368;  
Best Local Similarity 85.7%; Pred. No. 2.2e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WOSAFTK 7  
111111  
Db 169 WKSFTK 175

RESULT 13  
US-09-936-883A-21  
Sequence 21, Application US/09936883A  
GENERAL INFORMATION:  
APPLICANT: MIYATA, Toshio  
TITLE OF INVENTION: A Method for Detecting Megsin Protein and Use  
FILE REFERENCE: F2-101DIPCT  
CURRENT APPLICATION NUMBER: US/09/936,883A  
CURRENT FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: JP 1999-75305  
PRIOR FILING DATE: 1999-03-19  
PRIOR APPLICATION NUMBER: JP 1999-306623  
PRIOR FILING DATE: 1999-10-28  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: Patentln Ver. 2.0  
SEQ ID NO 21  
LENGTH: 368  
TYPE: PRT  
ORGANISM: Mus musculus  
US-09-936-883A-21

Query Match 90.0%; Score 36; DB 23; Length 368;  
Best Local Similarity 85.7%; Pred. No. 2.2e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WOSAFTK 7  
111111  
Db 169 WKSFTK 175

RESULT 14  
US-10-110-030-6  
Sequence 6, Application US/10110030  
GENERAL INFORMATION:  
APPLICANT: MIYATA, Toshio  
TITLE OF INVENTION: ANIMAL MODEL FOR MENENGEAL PROLIFERATIVE  
FILE REFERENCE: SHIM015  
CURRENT APPLICATION NUMBER: US/10/110,030  
CURRENT FILING DATE: 2002-04-03  
PRIOR APPLICATION NUMBER: 11/285736  
PRIOR FILING DATE: 1999-10-06  
NUMBER OF SEQ ID NOS: 15  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 6  
LENGTH: 368  
TYPE: PRT  
ORGANISM: Mus musculus  
US-10-110-030-6

Query Match 90.0%; Score 36; DB 25; Length 368;  
Best Local Similarity 85.7%; Pred. No. 2.2e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WOSAFTK 7  
111111  
Db 169 WKSFTK 175

RESULT 15  
US-09-508-997A-4  
Sequence 4, Application US/09508997A  
GENERAL INFORMATION:  
APPLICANT: MIYATA, Toshio  
TITLE OF INVENTION: Megsin Protein  
FILE REFERENCE: SHIM-004  
CURRENT APPLICATION NUMBER: US/09/508,997A  
CURRENT FILING DATE: 2000-06-27  
PRIOR APPLICATION NUMBER: JP 09-275302  
PRIOR FILING DATE: 1997-09-22  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: Patentln Ver. 2.0

SEQ ID NO 4  
 LENGTH: 380  
 TYPE: PRT  
 ORGANISM: Rattus rattus  
 FEATURE:  
 NAME/KEY: UNSURE  
 LOCATION: 51, 94  
 OTHER INFORMATION: Xaa is unknown.  
 US-09-508-997A-4

Query Match 90.0%; Score 36; DB 19; Length 380;  
 Best Local Similarity 85.7%; Pred. No. 2.3e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WQSAFTK 7  
 Db 181 WQSAFTK 187

Search completed: July 11, 2003, 12:27:28  
 Job time : 24.0333 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:55:27 ; Search time 6.57796 Seconds

(Without alignments)  
280.699 Million cell updates/sec

Title: US-10-091-442-2

Perfect score: 40

Sequence: 1 MOSAFTK 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1229476 seqs, 263775074 residues

Total number of hits satisfying chosen parameters: 1229476

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Pending\_Patents\_AA\_New:\*

1: /cgn2\_6/ptodata/1/paa/PCY\_NEW\_COMB.pep4:\*

2: /cgn2\_6/ptodata/1/paa/PCY\_NEW\_COMB.pep4:\*

3: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pep4:\*

4: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep4:\*

5: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep4:\*

6: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep4:\*

7: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep4:\*

8: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep4:\*

9: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep4:\*

10: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep4:\*

11: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep4:\*

12: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep4:\*

13: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep4:\*

14: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	100.0	380	10	US-09-936-883C-2
2	40	100.0	380	12	US-10-126-052A-263
3	40	100.0	380	14	US-60-443-566-3006
4	40	100.0	380	14	US-60-455-444-4692
5	40	100.0	380	14	US-60-465-241-4692
6	36	90.0	368	10	US-09-936-883C-21
7	36	90.0	380	10	US-09-936-883C-19
8	34	85.0	1717	10	US-09-958-359-20
9	33	82.5	60	12	US-10-424-599-183524
10	32	80.0	72	12	US-10-424-599-172361
11	32	80.0	369	12	US-10-369-493-12166
12	32	80.0	692	14	US-60-452-680-15528
13	32	80.0	938	12	US-10-282-122A-77952
14	32	80.0	938	12	US-10-282-122A-77952
15	32	80.0	944	12	US-10-289-762-500
16	32	80.0	944	10	US-09-949-016-6650
17	32	80.0	944	11	US-10-273-573-9504
18	32	80.0	944	12	US-10-218-140-6230
19	32	80.0	959	10	US-09-949-016-6904

20	32	80.0	959	12	US-10-219-051B-11182	Sequence 11182, A
21	32	80.0	981	10	US-09-949-016-11647	Sequence 11647, A
22	32	80.0	981	10	US-09-949-016-11648	Sequence 11648, A
23	32	80.0	981	12	US-10-087-192-624	Sequence 624, App
24	31	77.5	94	12	US-10-424-599-145438	Sequence 145438, A
25	31	77.5	93	12	US-10-424-599-153376	Sequence 153376, A
26	31	77.5	127	12	US-10-424-599-257209	Sequence 257209, A
27	31	77.5	121	12	US-10-219-382-1236	Sequence 1236, App
28	31	77.5	221	12	US-10-219-382A-1236	Sequence 1236, App
29	31	77.5	242	14	US-60-452-680-21389	Sequence 21389, A
30	31	77.5	242	14	US-60-470-166-1193	Sequence 1193, App
31	31	77.5	260	12	US-10-437-963-125687	Sequence 125687, A
32	31	77.5	264	12	US-10-289-762-486	Sequence 486, App
33	31	77.5	282	12	US-10-437-963-125689	Sequence 125689, A
34	31	77.5	514	12	US-10-369-493-21618	Sequence 21618, A
35	31	77.5	733	12	US-10-289-762-791	Sequence 791, App
36	31	77.5	978	10	US-09-841-260-65	Sequence 65, App1
37	31	77.5	978	12	US-10-312-573-35	Sequence 35, App1
38	31	77.5	1108	12	US-10-282-122A-60476	Sequence 60476, A
39	30	75.0	33	12	US-10-424-599-223477	Sequence 223477, A
40	30	75.0	75	12	US-10-424-599-204603	Sequence 204603, A
41	30	75.0	80	12	US-10-424-599-187032	Sequence 187032, A
42	30	75.0	88	10	US-09-864-408A-2936	Sequence 2936, App
43	30	75.0	91	10	US-09-864-408A-4728	Sequence 4728, App
44	30	75.0	130	2	PCT-US02-32727-15086	Sequence 15086, A
45	30	75.0	130	10	US-09-978-825-15086	Sequence 15086, A

#### ALIGNMENTS

RESULT 1  
US-09-936-883C-2  
Sequence 2, Application US/09936883C  
GENERAL INFORMATION:  
APPLICANT: MIYATA, Toshio  
TITLE OF INVENTION: A Method for Detecting Megsin Protein and Use  
FILE OF INVENTION: Theroft  
FILE REFERENCE: F2-101DIPCT  
CURRENT APPLICATION NUMBER: US/09/936,883C  
CURRENT FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: JP 1999-75305  
PRIOR FILING DATE: 1999-03-19  
PRIOR APPLICATION NUMBER: JP 1999-306623  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 380  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-936-883C-2

Query Match 100.0%; Score 40; DB 10; Length 380;  
Best Local Similarity 100.0%; Pred. No. 6.3;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOSAFTK 7  
DB 181 MOSAFTK 187

RESULT 2  
US-10-126-052A-263  
Sequence 263, Application US/10126052A  
GENERAL INFORMATION:  
APPLICANT: Aziz, Natasha  
APPLICANT: Murray, Richard  
APPLICANT: Eos Biotechnology, Inc.  
TITLE OF INVENTION: Methods of Diagnosis of Lung Cancer. Compositions and  
TITLE OF INVENTION: Methods of Screening for Modulators of Lung Cancer  
FILE REFERENCE: 018501-00153005  
CURRENT APPLICATION NUMBER: US/10/126,052A

```
;; CURRENT FILING DATE: 2002-04-18
;; PRIOR APPLICATION NUMBER: US 60/284,770
;; PRIOR FILING DATE: 2001-04-18
;; PRIOR APPLICATION NUMBER: US 60/290,492
;; PRIOR FILING DATE: 2001-05-10
;; PRIOR APPLICATION NUMBER: US 60/339,245
;; PRIOR FILING DATE: 2001-11-09
;; PRIOR APPLICATION NUMBER: US 60/350,666
;; PRIOR FILING DATE: 2001-11-13
;; PRIOR APPLICATION NUMBER: US 60/334,370
;; PRIOR FILING DATE: 2001-11-29
;; PRIOR APPLICATION NUMBER: US 60/372,246
;; PRIOR FILING DATE: 2002-04-12
;; NUMBER OF SEQ ID NOS: 691
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 263
;; LENGTH: 380
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-126-052A-263
```

```
Query Match
Best Local Similarity 100.0%; Score 40; DB 12; Length 380;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 WSAFTK 7
Db 181 WSAFTK 187
```

```
RESULT 3
US-60-443-566-3006
;; Sequence 3006, Application US/60443566
;; GENERAL INFORMATION:
;; APPLICANT: CARGILL, Michele
;; APPLICANT: BEGOVICH, Ann
;; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;; FILE REFERENCE: CLO01447
;; CURRENT APPLICATION NUMBER: US/60/443,566
;; CURRENT FILING DATE: 2003-01-30
;; NUMBER OF SEQ ID NOS: 25102
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 3006
;; LENGTH: 380
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-60-443-566-3006
```

```
Query Match
Best Local Similarity 100.0%; Score 40; DB 14; Length 380;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 WSAFTK 7
Db 181 WSAFTK 187
```

```
RESULT 4
US-60-455-444-4692
;; Sequence 4692, Application US/60455444
;; GENERAL INFORMATION:
;; APPLICANT: CARGILL, Michele
;; APPLICANT: BEGOVICH, Ann
;; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;; FILE REFERENCE: CLO01455
;; CURRENT APPLICATION NUMBER: US/60/455,444
;; CURRENT FILING DATE: 2003-03-18
;; NUMBER OF SEQ ID NOS: 50986
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 4692
;; LENGTH: 380
```

```
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-60-455-444-4692
```

```
Query Match
Best Local Similarity 100.0%; Score 40; DB 14; Length 380;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 WSAFTK 7
Db 181 WSAFTK 187
```

```
RESULT 5
US-60-465-241-4692
;; Sequence 4692, Application US/60465241
;; GENERAL INFORMATION:
;; APPLICANT: CARGILL, Michele
;; APPLICANT: BEGOVICH, Ann
;; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;; FILE REFERENCE: CLO01468
;; CURRENT APPLICATION NUMBER: US/60/465,241
;; CURRENT FILING DATE: 2003-04-23
;; NUMBER OF SEQ ID NOS: 258418
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 4692
;; LENGTH: 380
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-60-465-241-4692
```

```
Query Match
Best Local Similarity 100.0%; Score 40; DB 14; Length 380;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 WSAFTK 7
Db 181 WSAFTK 187
```

```
RESULT 6
US-09-936-883C-21
;; Sequence 21, Application US/09936883C
;; GENERAL INFORMATION:
;; APPLICANT: MIYATA, Toshio
;; TITLE OF INVENTION: A Method for Detecting Megsin Protein and Use
;; FILE REFERENCE: F2-101DIPCT
;; CURRENT APPLICATION NUMBER: US/09/936,883C
;; CURRENT FILING DATE: 2001-12-21
;; PRIOR APPLICATION NUMBER: JP 1999-75305
;; PRIOR FILING DATE: 1999-03-19
;; PRIOR APPLICATION NUMBER: JP 1999-306623
;; PRIOR FILING DATE: 1999-10-28
;; NUMBER OF SEQ ID NOS: 21
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 21
;; LENGTH: 368
;; TYPE: PRT
;; ORGANISM: Mus musculus
US-09-936-883C-21
```

```
Query Match
Best Local Similarity 90.0%; Score 36; DB 10; Length 368;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 WSAFTK 7
Db 169 WSAFTK 175
```

```
RESULT 7
```

US-09-936-883C-19  
: Sequence 19, Application US/09936883C  
: GENERAL INFORMATION:  
: APPLICANT: MIVATA, Toshio  
: TITLE OF INVENTION: A Method for Detecting Megsin Protein and Use  
: TITLE OF INVENTION: Theoret  
: FILE REFERENCE: P2-10IDPICT  
: CURRENT APPLICATION NUMBER: US/09/936, 883C  
: CURRENT FILING DATE: 2001-12-21  
: PRIOR APPLICATION NUMBER: JP 1999-75305  
: PRIOR FILING DATE: 1999-03-19  
: PRIOR APPLICATION NUMBER: JP 1999-306623  
: PRIOR FILING DATE: 1999-10-28  
: NUMBER OF SEQ ID NOS: 21  
: SOFTWARE: Patentln Ver. 2.0  
: SEQ ID NO 19  
: LENGTH: 380  
: TYPE: PRT  
: ORGANISM: Rattus norvegicus  
: FEATURE:  
: NAME/KEY: misc.feature  
: LOCATION: 51, 94  
: OTHER INFORMATION: Xaa is unknown.  
US-09-936-883C-19

Query Match 90.0%; Score 36; DB 10; Length 380;  
Best Local Similarity 85.7%; Pred. No. 40;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WOSAFTK 7  
|:|||||  
Db 181 WKSATK 187

RESULT 8  
US-09-958-359-20  
: Sequence 20, Application US/09958359  
: GENERAL INFORMATION:  
: APPLICANT: SUGEN, INC.  
: TITLE OF INVENTION: TYROSINE KINASE SUBSTRATE (TKS) PROTEINS  
: FILE REFERENCE: 038602/0102  
: CURRENT APPLICATION NUMBER: US/09/958, 359  
: CURRENT FILING DATE: 2002-02-05  
: NUMBER OF SEQ ID NOS: 55  
: SOFTWARE: Patentln Ver. 2.1  
: SEQ ID NO 20  
: LENGTH: 1717  
: TYPE: PRT  
: ORGANISM: Unknown Organism  
: FEATURE:  
: OTHER INFORMATION: Description of Unknown Organism: Grub  
US-09-958-359-20

Query Match 85.0%; Score 34; DB 10; Length 1717;  
Best Local Similarity 85.7%; Pred. No. 3.8e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 WOSAFTK 7  
|:|||||  
Db 467 WOSALTK 473

RESULT 9  
US-10-424-599-183524  
: Sequence 183524, Application US/10424599  
: GENERAL INFORMATION:  
: APPLICANT: La Rosa Thomas J  
: APPLICANT: Kovalic David K  
: APPLICANT: Zhou Yihua  
: APPLICANT: Cao Yongwei  
: TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
: TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
: FILE REFERENCE: 38-21(53223)B

: CURRENT APPLICATION NUMBER: US/10/424, 599  
: CURRENT FILING DATE: 2003-04-28  
: NUMBER OF SEQ ID NOS: 285684  
: SEQ ID NO 183524  
: LENGTH: 60  
: TYPE: PRT  
: ORGANISM: Glycine max  
: FEATURE:  
: NAME/KEY: unsure  
: LOCATION: (1)..(60)  
: OTHER INFORMATION: unsure at all Xaa locations  
: FEATURE:  
: OTHER INFORMATION: Clone ID: PAT\_MRT3847\_136735C.1.pep  
US-10-424-599-183524

Query Match 82.5%; Score 33; DB 12; Length 60;  
Best Local Similarity 71.4%; Pred. No. 32;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 WOSAFTK 7  
|:|||||  
Db 19 WOSAYTE 25

RESULT 10  
US-10-424-599-172361  
: Sequence 172361, Application US/10424599  
: GENERAL INFORMATION:  
: APPLICANT: La Rosa Thomas J  
: APPLICANT: Kovalic David K  
: APPLICANT: Zhou Yihua  
: APPLICANT: Cao Yongwei  
: TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
: TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
: FILE REFERENCE: 38-21(53223)B  
: CURRENT APPLICATION NUMBER: US/10/424, 599  
: CURRENT FILING DATE: 2003-04-28  
: NUMBER OF SEQ ID NOS: 285684  
: SEQ ID NO 172361  
: LENGTH: 72  
: TYPE: PRT  
: ORGANISM: Glycine max  
: FEATURE:  
: OTHER INFORMATION: Clone ID: PAT\_MRT3847\_126659C.1.pep  
US-10-424-599-172361

Query Match 80.0%; Score 32; DB 12; Length 72;  
Best Local Similarity 83.3%; Pred. No. 60;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WOSAFT 6  
|:|||||  
Db 5 WOSST 10

RESULT 11  
US-10-369-493-12166  
: Sequence 12166, Application US/10369493  
: GENERAL INFORMATION:  
: APPLICANT: Cao, Yongwei  
: APPLICANT: Hinkle, Gregory J.  
: APPLICANT: Slater, Steven C.  
: APPLICANT: Goldman, Barry S.  
: APPLICANT: Chen, Xianfeng  
: TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
: TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
: FILE REFERENCE: 38-10(52052)B  
: CURRENT APPLICATION NUMBER: US/10/369, 493  
: CURRENT FILING DATE: 2003-02-28  
: PRIOR APPLICATION NUMBER: US 60/360, 039  
: PRIOR FILING DATE: 2002-02-21  
: NUMBER OF SEQ ID NOS: 47374  
: SEQ ID NO 12166

```
; LENGTH: 369
; TYPE: PRT
; ORGANISM: Mesorhizobium loti
US-10-369-493-12166

Query Match
Best Local Similarity 80.0%; Score 32; DB 12; Length 369;
                          71.4%; Pred. No. 2.5e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

OY 1 WOSAFTK 7  
111111  
Db 315 WOSAFTK 321

RESULT 12  
US-60-452-680-15528  
; Sequence 15528, Application US/60452680  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele  
; APPLICANT: GRUPE, Andrew  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; FILE REFERENCE: C1001450  
; CURRENT APPLICATION NUMBER: US/60/452,680  
; CURRENT FILING DATE: 2003-03-07  
; NUMBER OF SEQ ID NOS: 116213  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15528  
; LENGTH: 692  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-60-452-680-15528

Query Match  
Best Local Similarity 80.0%; Score 32; DB 14; Length 692;  
 71.4%; Pred. No. 4.3e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 WOSAFTK 7  
111111  
Db 460 WSSSFTK 466

RESULT 13  
US-10-282-122A-77952  
; Sequence 77952, Application US/10282122A  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA 034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578

```
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 77952
; LENGTH: 837
; TYPE: PRT
; ORGANISM: Yersinia pestis
US-10-282-122A-77952

Query Match
Best Local Similarity 80.0%; Score 32; DB 12; Length 837;
                          83.3%; Pred. No. 5.1e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

OY 1 WOSAFT 6  
111111  
Db 220 WOSAFT 225

RESULT 14  
US-10-312-273-5  
; Sequence 5, Application US/10312273  
; GENERAL INFORMATION:  
; APPLICANT: CHIRON SPA  
; TITLE OF INVENTION: IMMUNISATION AGAINST CHLAMYDIA PNEUMONIAE  
; FILE REFERENCE: P025035MO  
; CURRENT APPLICATION NUMBER: US/10/312,273  
; CURRENT FILING DATE: 2002-12-20  
; PRIOR APPLICATION NUMBER: 0016363.4  
; PRIOR FILING DATE: 2000-07-03  
; PRIOR APPLICATION NUMBER: 0017047.2  
; PRIOR FILING DATE: 2000-07-11  
; PRIOR APPLICATION NUMBER: 0017983.8  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: 0019368.0  
; PRIOR FILING DATE: 2000-08-07  
; PRIOR APPLICATION NUMBER: 0020440.4  
; PRIOR FILING DATE: 2000-08-18  
; PRIOR APPLICATION NUMBER: 0022583.9  
; PRIOR FILING DATE: 2000-09-14  
; PRIOR APPLICATION NUMBER: 0027549.5  
; PRIOR FILING DATE: 2000-11-10  
; PRIOR APPLICATION NUMBER: 0031706.5  
; PRIOR FILING DATE: 2000-12-22  
; NUMBER OF SEQ ID NOS: 664  
; SOFTWARE: Seqwin99, version 1.02  
; SEQ ID NO 5  
; LENGTH: 938  
; TYPE: PRT  
; ORGANISM: Chlamydia pneumoniae  
US-10-312-273-5

Query Match  
Best Local Similarity 80.0%; Score 32; DB 12; Length 938;  
 83.3%; Pred. No. 5.6e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WOSAFT 6  
111111  
Db 642 WOSAFT 647

RESULT 15  
US-10-289-762-500  
; Sequence 500, Application US/10289762  
; GENERAL INFORMATION:

; APPLICANT: Griffiths, R.  
 ; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments  
 ; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention  
 ; TITLE OF INVENTION: and treatment of infection  
 ; FILE REFERENCE: 9710-003-999  
 ; CURRENT APPLICATION NUMBER: US/10/289,762  
 ; CURRENT FILING DATE: 2003-03-27  
 ; NUMBER OF SEQ ID NOS: 6849  
 ; SEQ ID NO 500  
 ; LENGTH: 940  
 ; TYPE: PRT  
 ; ORGANISM: Chlamydia pneumoniae  
 US-10-289-762-500

Query Match 80.0% Score 32; DB 12; Length 940;  
 Best Local Similarity 83.3%; Pred. NO. 5.7e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WOSAFT 6  
 ||||:|  
 Db 644 WOSAFT 649

Search completed: July 11, 2003, 12:35:05  
 Job time : 7.67796 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: July 11, 2003, 11:54:32 ; Search time 59.8046 Seconds

(without alignments)  
204.833 Million cell updates/sec

Title: US-10-091-442-3

Perfect score: 100

Sequence: 1 VERVDFTHLEDTRRNINK 19

## Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 64473110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

Pending\_Patents\_AA\_Main:\*  
1: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep:\*  
7: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep:\*  
8: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep:\*  
9: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep:\*  
10: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep:\*  
11: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep:\*  
12: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep:\*  
13: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep:\*  
14: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep:\*  
15: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep:\*  
16: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep:\*  
17: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep:\*  
18: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep:\*  
19: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep:\*  
20: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep:\*  
21: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep:\*  
22: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep:\*  
23: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep:\*  
24: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep:\*  
25: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep:\*  
26: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep:\*  
27: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	100	100.0	19	US-09-140-719-3	Sequence 3, Appli
2	100	100.0	19	US-10-091-442-3	Sequence 3, Appli
3	100	100.0	41	US-60-196-718-3839	Sequence 3839, Ap
4	100	100.0	380	US-08-091-028A-34	Sequence 34, Appli
5	100	100.0	380	US-09-140-719-34	Sequence 34, Appli
6	100	100.0	380	US-09-508-997A-2	Sequence 2, Appli

7	100	100.0	380	US-09-791-537-281	Sequence 281, App
8	100	100.0	380	US-09-936-883A-2	Sequence 2, Appli
9	100	100.0	380	US-10-091-442-34	Sequence 34, Appli
10	100	100.0	380	US-10-110-030-2	Sequence 2, Appli
11	82	82.0	19	US-08-091-028-3	Sequence 3, Appli
12	82	82.0	19	US-08-091-028A-3	Sequence 3, Appli
13	71	71.0	368	US-09-508-997A-6	Sequence 6, Appli
14	71	71.0	368	US-09-936-883A-21	Sequence 21, Appli
15	71	71.0	368	US-10-110-030-6	Sequence 6, Appli
16	68	68.0	380	US-09-508-997A-4	Sequence 4, Appli
17	68	68.0	380	US-09-936-883A-19	Sequence 19, Appli
18	68	68.0	380	US-10-110-030-4	Sequence 4, Appli
19	51.5	51.5	509	US-09-791-537-14188	Sequence 14188, A
20	51.5	51.5	509	US-09-791-537-88278	Sequence 88278, A
21	50	50.0	338	US-60-389-987-1574	Sequence 1574, Ap
22	50	50.0	338	US-60-412-418-1574	Sequence 1574, Ap
23	50	50.0	390	PCT-US02-07215-12	Sequence 12, Appli
24	50	50.0	390	US-08-099-259-2	Sequence 2, Appli
25	50	50.0	390	US-08-568-147A-2	Sequence 2, Appli
26	50	50.0	390	US-08-731-566-2	Sequence 2, Appli
27	50	50.0	390	US-09-653-464B-9	Sequence 9, Appli
28	50	50.0	390	US-09-791-537-33386	Sequence 33386, A
29	50	50.0	390	US-09-791-537-76123	Sequence 76123, A
30	50	50.0	390	US-09-791-537-121576	Sequence 121576, A
31	50	50.0	390	US-09-823-187-29	Sequence 29, Appli
32	50	50.0	390	US-09-823-187-30	Sequence 30, Appli
33	50	50.0	390	US-10-037-417-65	Sequence 65, Appli
34	50	50.0	390	US-10-094-944-12	Sequence 12, Appli
35	50	50.0	413	US-09-760-492-106	Sequence 106, App
36	50	50.0	413	US-10-162-749-106	Sequence 106, App
37	49	49.0	226	US-09-252-991A-26823	Sequence 26823, A
38	48.5	48.5	85	US-09-758-448-934	Sequence 934, App
39	48.5	48.5	85	US-10-212-760-934	Sequence 934, App
40	48.5	48.5	163	US-09-758-448-304	Sequence 304, App
41	48.5	48.5	163	US-10-212-215-304	Sequence 304, App
42	48	48.0	480	US-10-219-999-62273	Sequence 62273, A
43	48	48.0	480	US-60-324-109-28371	Sequence 28371, A
44	48	48.0	688	US-10-219-999-32900	Sequence 32900, A
45	48	48.0	688	US-60-324-109-18095	Sequence 18095, A

## ALIGNMENTS

RESULT 1  
US-09-140-719-3  
; Sequence 3, Application US/09140719  
; GENERAL INFORMATION:  
; APPLICANT: TSUJIMOTO, Masafumi  
; APPLICANT: IWASA, Fuyuki  
; APPLICANT: TSUROOKA, Nobuo  
; APPLICANT: NAKAZATO, Hiroshi  
; APPLICANT: MIURA, Kenju  
; APPLICANT: ISHIDA, Nobuhiko  
; APPLICANT: KURIHARA, Tatsuya  
; APPLICANT: YAMAICHI, Kozo  
; APPLICANT: YAMAGUCHI, Nozomi  
; TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
; NUMBER OF SEQUENCES: 34  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: P.O. Box 1404  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/140,719

FILING DATE: 08-AUG-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-140-719-3

Query Match 100.0%; Score 100; DB 15; Length 19;  
Best Local Similarity 100.0%; Pred. No. 2.2e-09;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VERVDFTNHLDETRRNINK 19  
Db 1 VERVDFTNHLDETRRNINK 19

## RESULT 2

US-10-091-442-3

Sequence 34, Application US/10091442

GENERAL INFORMATION:

APPLICANT: TSUJIMOTO, Masafumi

IMASA, Fuyuki

TSURUOKA, Nobuo

NAKAZATO, Hiroshi

MURA, Kenji

ISHIDA, Nobuhiko

KURIHARA, Tatsuya

YAMAICHI, Kozo

YAMAGUCHI, Nozomi

TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR

NUMBER OF SEQUENCES: 34

CORRESPONDENCE ADDRESS:

ADDRESS: Burns, Doane, Swecker &amp; Mathis

STREET: P.O. Box 1404

CITY: Alexandria

STATE: Virginia

COUNTRY: United States

ZIP: 22313-1404

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/091,442

FILING DATE: 07-Mar-2002

CLASSIFICATION: &lt;Unknown&gt;

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/140,719

FILING DATE: 08-AUG-1998

APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-10-091-442-3

Query Match 100.0%; Score 100; DB 24; Length 19;  
Best Local Similarity 100.0%; Pred. No. 2.2e-09;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VERVDFTNHLDETRRNINK 19  
Db 1 VERVDFTNHLDETRRNINK 19

## RESULT 3

US-60-196-718-3839

Sequence 3839, Application US/60196718

GENERAL INFORMATION:

APPLICANT: Bonazzi, Vivien

TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,

TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND

FILE REFERENCE: CL000456

CURRENT APPLICATION NUMBER: US/60/196,718

CURRENT FILING DATE: 2000-04-13

NUMBER OF SEQ ID NOS: 7494

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 3839

LENGTH: 41

TYPE: PRT

ORGANISM: HUMAN

FEATURE:

NAME/KEY: VARIANT

LOCATION: (1)...(41)

OTHER INFORMATION: Xaa = Any Amino Acid

US-60-196-718-3839

Query Match 100.0%; Score 100; DB 27; Length 41;  
Best Local Similarity 100.0%; Pred. No. 5.7e-09;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VERVDFTNHLDETRRNINK 19  
Db 15 VERVDFTNHLDETRRNINK 33

## RESULT 4

US-08-091-028A-34

Sequence 34, Application US/08091028A

GENERAL INFORMATION:

APPLICANT: TSUJIMOTO, Masafumi

IMASA, Fuyuki

TSURUOKA, Nobuo



APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, Nobuhiko  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAUCHI, Kozo  
APPLICANT: YAMAGUCHI, Nozomi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/091,028A  
FILING DATE: 14-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: REA, TERESA STANER  
REGISTRATION NUMBER: 30,427  
REFERENCE/DOCKET NUMBER: 001560-204  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-6620  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-091-028A-34

Query Match 100.0%; Score 100; DB 4; Length 380;  
Best Local Similarity 100.0%; Pred. No. 9.1e-08;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VERVDFTNHLDTNRNINK 19  
DB 126 VERVDFTNHLDTNRNINK 144

RESULT 5  
US-09-140-719-34  
Sequence 34, Application US/09140719  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masaaki  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSUROOKA, Nobuo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, Nobuhiko  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAUCHI, Kozo  
APPLICANT: YAMAGUCHI, Nozomi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria

STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-140-719-34

Query Match 100.0%; Score 100; DB 15; Length 380;  
Best Local Similarity 100.0%; Pred. No. 9.1e-08;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VERVDFTNHLDTNRNINK 19  
DB 126 VERVDFTNHLDTNRNINK 144

RESULT 6  
US-09-508-997A-2  
Sequence 2, Application US/09508997A  
GENERAL INFORMATION:  
APPLICANT: MIYATA, Toshio  
TITLE OF INVENTION: Megsin Protein  
FILE REFERENCE: SHIM-004  
CURRENT APPLICATION NUMBER: US/09/508,997A  
CURRENT FILING DATE: 2000-06-27  
PRIOR APPLICATION NUMBER: JP 09-275302  
PRIOR FILING DATE: 1997-09-22  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: Patent Ver. 2.0  
SEQ ID NO 2  
LENGTH: 380  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-508-997A-2

Query Match 100.0%; Score 100; DB 19; Length 380;  
Best Local Similarity 100.0%; Pred. No. 9.1e-08;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VERVDFTNHLDTNRNINK 19  
DB 126 VERVDFTNHLDTNRNINK 144

Db 126 VERVDFTNHLDETRRNINK 144

## RESULT 7

US-09-791-537-281  
; Sequence 281, Application US/09791537  
; GENERAL INFORMATION:  
; APPLICANT: Biomolix, Inc.  
; APPLICANT: Debe, Derek  
; APPLICANT: Danzer, Joseph  
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB  
; FILE REFERENCE: 261/210  
; CURRENT APPLICATION NUMBER: US/09/791,537  
; CURRENT FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 153055  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 281  
; LENGTH: 380  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-791-537-281

Query Match 100.0%; Score 100; DB 21; Length 380;  
Best Local Similarity 100.0%; Pred. No. 9, 1e-08;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VERVDFTNHLDETRRNINK 19  
Db 126 VERVDFTNHLDETRRNINK 144

## RESULT 8

US-09-936-883A-2  
; Sequence 2, Application US/09936883A  
; GENERAL INFORMATION:  
; APPLICANT: MIYATA, Toshio  
; TITLE OF INVENTION: A Method for Detecting Magsin Protein and Use  
; FILE REFERENCE: P2-101DPICT  
; CURRENT APPLICATION NUMBER: US/09/936,883A  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: JP 1999-75305  
; PRIOR FILING DATE: 1999-03-19  
; PRIOR APPLICATION NUMBER: JP 1999-306623  
; PRIOR FILING DATE: 1999-10-28  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 380  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-936-883A-2

Query Match 100.0%; Score 100; DB 23; Length 380;  
Best Local Similarity 100.0%; Pred. No. 9, 1e-08;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VERVDFTNHLDETRRNINK 19  
Db 126 VERVDFTNHLDETRRNINK 144

## RESULT 9

US-10-091-442-34  
; Sequence 34, Application US/10091442  
; GENERAL INFORMATION:  
; APPLICANT: TSUJIMOTO, Masafumi  
; APPLICANT: IWASA, Fuyuki  
; APPLICANT: TSUROOKA, Nobuo  
; APPLICANT: NAKAZATO, Hiroshi  
; APPLICANT: MIURA, Kenju  
; APPLICANT: ISHIDA, Nobuhito

KURIHARA, Tatsuya  
YAMAICHI, Koza  
YAMAGUCHI, Nozomi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
City: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404

COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/091,442  
FILING DATE: 07-Mar-2002  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021

INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 34:  
US-10-091-442-34

Query Match 100.0%; Score 100; DB 24; Length 380;  
Best Local Similarity 100.0%; Pred. No. 9, 1e-08;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VERVDFTNHLDETRRNINK 19  
Db 126 VERVDFTNHLDETRRNINK 144

## RESULT 10

US-10-110-030-2  
; Sequence 2, Application US/10110030  
; GENERAL INFORMATION:  
; APPLICANT: Miyata, Toshio  
; TITLE OF INVENTION: ANIMAL MODEL FOR MESENTERIAL PROLIFERATIVE  
; FILE REFERENCE: SHIM015  
; CURRENT APPLICATION NUMBER: US/10/110,030  
; CURRENT FILING DATE: 2002-04-03  
; PRIOR APPLICATION NUMBER: 11/285736  
; PRIOR FILING DATE: 1999-10-06  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 380

TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-110-030-2

Query Match 100.0%; Score 100; DB 25; Length 380;  
Best Local Similarity 100.0%; Pred. No. 9.1e-08;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VERVDFTNHLDTNRNIN 19  
DB 126 VERVDFTNHLDTNRNIN 144

RESULT 11  
US-08-091-028-3  
Sequence 3, Application US/08091028  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSURUOKA, Nobuo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, Nobuhiko  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAUCHI, Kozo  
APPLICANT: YAMAGUCHI, Nozomi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/091.028  
FILING DATE: 14-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Crane-Feury, Sharon E  
REGISTRATION NUMBER: 36,113  
REFERENCE/DOCKET NUMBER: 001560-204  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-6620  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-091-028-3

Query Match 82.0%; Score 82; DB 4; Length 19;  
Best Local Similarity 84.2%; Pred. No. 2.3e-06;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 VERVDFTNHLDTNRNIN 19  
DB 1 VEXVDFTNHLDTXNIN 19

RESULT 12  
US-08-091-028A-3  
Sequence 3, Application US/08091028A  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSURUOKA, Nobuo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, Nobuhiko  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAUCHI, Kozo  
APPLICANT: YAMAGUCHI, Nozomi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/091.028A  
FILING DATE: 14-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: REA, TERESA STANER  
REGISTRATION NUMBER: 30,427  
REFERENCE/DOCKET NUMBER: 001560-204  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-6620  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-091-028A-3

Query Match 82.0%; Score 82; DB 4; Length 19;  
Best Local Similarity 84.2%; Pred. No. 2.3e-06;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 VERVDFTNHLDTNRNIN 19  
DB 1 VEXVDFTNHLDTXNIN 19

RESULT 13  
US-09-508-997A-6  
Sequence 6, Application US/09508997A  
GENERAL INFORMATION:  
APPLICANT: MIYATA, Toshio  
TITLE OF INVENTION: Megsin Protein  
FILE REFERENCE: SHIM-004  
CURRENT APPLICATION NUMBER: US/09/508.997A  
CURRENT FILING DATE: 2000-06-27

; PRIOR APPLICATION NUMBER: JP 09-275302  
; PRIOR FILING DATE: 1997-09-22  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: Patentln Ver. 2.0  
; SEQ ID NO: 6  
; LENGTH: 368  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-508-997A-6

Query Match 71.0%; Score 71; DB 19; Length 368;  
Best Local Similarity 73.7%; Pred. No. 0.0065;  
Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 VERVDFTNHLEDTRRNK 19  
|||||:|||||  
Db 114 VERVDFTNVDTRFRK 132

RESULT 14  
US-09-936-883A-21  
; Sequence 21, Application US/09936883A  
; GENERAL INFORMATION:  
; APPLICANT: MIYATA, Toshio  
; TITLE OF INVENTION: A Method for Detecting Megsin Protein and Use  
; FILE REFERENCE: P2-10IDPICT  
; CURRENT APPLICATION NUMBER: US/09/936, 883A  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: JP 1999-75305  
; PRIOR FILING DATE: 1999-03-19  
; PRIOR APPLICATION NUMBER: JP 1999-306623  
; PRIOR FILING DATE: 1999-10-28  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: Patentln Ver. 2.0  
; SEQ ID NO: 21  
; LENGTH: 368  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-936-883A-21

Query Match 71.0%; Score 71; DB 23; Length 368;  
Best Local Similarity 73.7%; Pred. No. 0.0065;  
Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 VERVDFTNHLEDTRRNK 19  
|||||:|||||  
Db 114 VERVDFTNVDTRFRK 132

RESULT 15  
US-10-110-030-6  
; Sequence 6, Application US/10110030  
; GENERAL INFORMATION:  
; APPLICANT: Miyata, Toshio  
; TITLE OF INVENTION: ANIMAL MODEL FOR MESENIAL PROLIFERATIVE  
; FILE REFERENCE: SHIM015  
; CURRENT APPLICATION NUMBER: US/10/110, 030  
; PRIOR FILING DATE: 2002-04-03  
; PRIOR APPLICATION NUMBER: 11/285736  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO: 6  
; LENGTH: 368  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-10-110-030-6

Query Match 71.0%; Score 71; DB 25; Length 368;  
Best Local Similarity 73.7%; Pred. No. 0.0065;  
Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 VERVDFTNHLEDTRRNK 19  
|||||:|||||  
Db 114 VERVDFTNVDTRFRK 132

Search completed: July 11, 2003, 12:27:30  
Job time : 61.8046 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: July 11, 2003, 11:55:27 : Search time 17.8545 Seconds

(without alignments)  
280.699 Million cell updates/sec

Title: US-10-091-442-3

Perfect score: 100

Sequence: 1 VERVDFNHEDETRRNINK 19

## Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1229476 seqs, 263775074 residues

Total number of hits satisfying chosen parameters: 1229476

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

```
1: /cgn2_6/ptodata/1/paa/PCR_NEW_COMB.pep4:*
2: /cgn2_6/ptodata/1/paa/PCR_NEW_COMB.pep4:*
3: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep4:*
4: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep4:*
5: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep4:*
6: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep4:*
7: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep4:*
8: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep4:*
9: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep4:*
10: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep4:*
11: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep4:*
12: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep4:*
13: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pep4:*
14: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pep4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	100	100.0	380	US-09-936-883C-2	Sequence 2, Appl1
2	100	100.0	380	US-10-126-052A-263	Sequence 263, App
3	100	100.0	380	US-60-443-566-3006	Sequence 3006, App
4	100	100.0	380	US-60-455-444-4692	Sequence 4692, App
5	100	100.0	380	US-60-465-241-4692	Sequence 4692, App
6	71	71.0	368	US-09-936-883C-21	Sequence 21, Appl
7	68	68.0	380	US-09-936-883C-19	Sequence 19, Appl
8	50	50.0	328	US-60-443-566-4373	Sequence 4373, App
9	50	50.0	328	US-60-455-444-7455	Sequence 7455, App
10	50	50.0	328	US-60-465-241-7455	Sequence 7455, App
11	50	50.0	390	PCT-US02-072158-12	Sequence 12, Appl
12	50	50.0	390	PCT-US03-17409-175	Sequence 175, App
13	50	50.0	390	US-60-448-285-5	Sequence 5, Appl1
14	50	50.0	410	US-09-724-676-57391	Sequence 57391, A
15	50	50.0	410	US-09-724-676A-57391	Sequence 57391, A
16	49	49.0	226	US-10-366-683-26823	Sequence 26823, A
17	49	49.0	226	US-10-419-128-26823	Sequence 26823, A
18	49	49.0	1169	US-10-282-122A-47218	Sequence 47218, A
19	48.5	48.5	235	US-09-949-016-11482	Sequence 11482, A

20	48	48.0	481	US-10-425-114-68385	Sequence 68385, A
21	48	48.0	689	US-10-425-114-45472	Sequence 45472, A
22	47	47.0	503	PCT-US02-20350-4	Sequence 4, Appl1
23	47	47.0	503	US-10-219-051B-14336	Sequence 14336, A
24	47	47.0	504	PCT-US02-20350-6	Sequence 6, Appl1
25	47	47.0	504	US-10-219-051B-14338	Sequence 14338, A
26	47	47.0	504	US-60-452-680-24158	Sequence 24158, A
27	47	47.0	504	US-60-453-135-15050	Sequence 15050, A
28	47	47.0	504	US-60-453-050-15050	Sequence 15050, A
29	47	47.0	504	US-60-466-412-15050	Sequence 15050, A
30	47	47.0	506	US-10-373-877-18	Sequence 18, Appl
31	47	47.0	507	US-09-949-016-10802	Sequence 10802, A
32	47	47.0	1043	US-09-949-016-11650	Sequence 11650, A
33	46	46.0	215	US-60-443-566-4034	Sequence 4034, App
34	46	46.0	215	US-60-455-444-5240	Sequence 5240, App
35	46	46.0	215	US-60-465-241-5240	Sequence 5240, App
36	46	46.0	258	US-60-443-566-4033	Sequence 4033, App
37	46	46.0	258	US-60-455-444-5239	Sequence 5239, App
38	46	46.0	258	US-60-465-241-5239	Sequence 5239, App
39	46	46.0	375	PCT-US02-19592-93	Sequence 93, Appl
40	46	46.0	375	PCT-US03-17409-189	Sequence 189, App
41	46	46.0	375	US-10-126-052A-36	Sequence 36, Appl
42	46	46.0	375	US-10-376-564-3	Sequence 3, Appl1
43	46	46.0	375	US-10-376-564-4	Sequence 4, Appl1
44	46	46.0	375	US-60-438-735-166	Sequence 166, App
45	46	46.0	375	US-60-449-629-149	Sequence 149, App

## ALIGNMENTS

```
RESULT 1
US-09-936-883C-2
; Sequence 2, Application US/09936883C
; GENERAL INFORMATION:
; APPLICANT: MIVATA, Toshio
; TITLE OF INVENTION: A Method for Detecting Megsin Protein and Use
; FILE REFERENCE: F2-101DIPCT
; CURRENT APPLICATION NUMBER: US/09/936, 883C
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: JP 1999-75305
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: JP 1999-306623
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-936-883C-2

Query Match      100.0%; Score 100; DB 10; Length 380;
Best Local Similarity 100.0%; Pred. No. 7.2e-08;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 VERVDFNHEDETRRNINK 19
Db      126 VERVDFNHEDETRRNINK 144

RESULT 2
US-10-126-052A-263
; Sequence 263, Application US/10126052A
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Murray, Richard
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Lung Cancer, Compositions and
; TITLE OF INVENTION: Methods of Screening for Modulators of Lung Cancer
; FILE REFERENCE: 018501-001530US
; CURRENT APPLICATION NUMBER: US/10/126, 052A
```

```
; CURRENT FILING DATE: 2002-04-18
; PRIOR APPLICATION NUMBER: US 60/284,770
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: US 60/290,492
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: US 60/339,245
; PRIOR FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/334,370
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/372,246
; PRIOR FILING DATE: 2002-04-12
; NUMBER OF SEQ ID NOS: 691
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 263
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-126-052A-263
```

```
Query Match          100.0%; Score 100; DB 12; Length 380;
Best Local Similarity 100.0%; Pred. No. 7.2e+08;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 VERVDFTNHLDETRRINK 19
        |||
Db      126 VERVDFTNHLDETRRINK 144
```

```
RESULT 3
US-60-443-566-3006
; Sequence 3006, Application US/60443566
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001447
; CURRENT APPLICATION NUMBER: US/60/443,566
; CURRENT FILING DATE: 2003-01-30
; NUMBER OF SEQ ID NOS: 25102
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3006
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-443-566-3006
```

```
Query Match          100.0%; Score 100; DB 14; Length 380;
Best Local Similarity 100.0%; Pred. No. 7.2e+08;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 VERVDFTNHLDETRRINK 19
        |||
Db      126 VERVDFTNHLDETRRINK 144
```

```
RESULT 4
US-60-455-444-4692
; Sequence 4692, Application US/60455444
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: BEGOVICH, Ann
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001453
; CURRENT APPLICATION NUMBER: US/60/455,444
; CURRENT FILING DATE: 2003-03-18
; NUMBER OF SEQ ID NOS: 50986
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4692
; LENGTH: 380
```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-455-444-4692
```

```
Query Match          100.0%; Score 100; DB 14; Length 380;
Best Local Similarity 100.0%; Pred. No. 7.2e+08;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 VERVDFTNHLDETRRINK 19
        |||
Db      126 VERVDFTNHLDETRRINK 144
```

```
RESULT 5
US-60-465-241-4692
; Sequence 4692, Application US/60465241
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: BEGOVICH, Ann
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001468
; CURRENT APPLICATION NUMBER: US/60/465,241
; CURRENT FILING DATE: 2003-04-23
; NUMBER OF SEQ ID NOS: 258418
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4692
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-465-241-4692
```

```
Query Match          100.0%; Score 100; DB 14; Length 380;
Best Local Similarity 100.0%; Pred. No. 7.2e+08;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 VERVDFTNHLDETRRINK 19
        |||
Db      126 VERVDFTNHLDETRRINK 144
```

```
RESULT 6
US-09-936-883C-21
; Sequence 21, Application US/09936883C
; GENERAL INFORMATION:
; APPLICANT: MIYATA, Toshio
; TITLE OF INVENTION: A Method for Detecting Megsin Protein and Use
; FILE REFERENCE: P2-101D1PCT
; CURRENT APPLICATION NUMBER: US/09/936,883C
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: JP 1999-75305
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: JP 1999-306623
; PRIOR FILING DATE: 1999-10-28
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 368
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-936-883C-21
```

```
Query Match          71.0%; Score 71; DB 10; Length 368;
Best Local Similarity 73.7%; Pred. No. 0.0034;
Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      1 VERVDFTNHLDETRRINK 19
        |||
Db      114 VERVDFTNVDQDTRRINK 132
```

```
RESULT 7
```

```

US-09-936-883C-19
: Sequence 19, Application US/09936883C
: GENERAL INFORMATION:
: APPLICANT: MIYATA, Toshio
: TITLE OF INVENTION: A method for Detecting Mesgin Protein and Use
: TITLE OF INVENTION: Theorect
: FILE REFERENCE: F2-101DPLPCT
: CURRENT APPLICATION NUMBER: US/09/936,883C
: CURRENT FILING DATE: 2001-12-21
: PRIOR APPLICATION NUMBER: JP 1999-75305
: PRIOR FILING DATE: 1999-03-19
: PRIOR APPLICATION NUMBER: JP 1999-306623
: PRIOR FILING DATE: 1999-10-28
: NUMBER OF SEQ ID NOS: 21
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 19
: LENGTH: 380
: TYPE: PRT
: ORGANISM: Rattus norvegicus
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 51, 94
: OTHER INFORMATION: Xaa is unknown.
US-09-936-883C-19

Query Match          68.0%  Score 68;  DB 10;  Length 380;
Best Local Similarity 68.4%  Pred. No. 0.011;
Matches 13;  Conservative 3;  Mismatches 3;  Indels 0;  Gaps 0;

OY      1  VERVDFTNHLIEDTRRNINK 19
        |||||:::|||||
Db      126  VERVDFTNDIQETRRKINK 144

RESULT 8
US-60-443-566-4373
: Sequence 4373, Application US/60443566
: GENERAL INFORMATION:
: APPLICANT: CARGILL, Michele
: APPLICANT: BEGOVICH, Ann
: TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
: TITLE OF INVENTION: RHUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
: FILE REFERENCE: CL001447
: CURRENT APPLICATION NUMBER: US/60/443,566
: CURRENT FILING DATE: 2003-01-30
: NUMBER OF SEQ ID NOS: 25102
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 4373
: LENGTH: 328
: TYPE: PRT
: ORGANISM: Homo sapiens
US-60-443-566-4373

Query Match          50.0%  Score 50;  DB 14;  Length 328;
Best Local Similarity 55.6%  Pred. No. 7.4;
Matches 10;  Conservative 3;  Mismatches 5;  Indels 0;  Gaps 0;

OY      1  VERVDFTNHLIEDTRRNIN 18
        |||||:::|||||
Db      69  VESVDYFANAPESRSKKIN 86

RESULT 9
US-60-455-444-7455
: Sequence 7455, Application US/60455444
: GENERAL INFORMATION:
: APPLICANT: CARGILL, Michele
: APPLICANT: BEGOVICH, Ann
: TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
: TITLE OF INVENTION: RHUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
: FILE REFERENCE: CL001455
: CURRENT APPLICATION NUMBER: US/60/455,444
: CURRENT FILING DATE: 2003-03-18

```

```

: NUMBER OF SEQ. ID NOS: 50986
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 7455
: LENGTH: 328
: TYPE: PRT
: ORGANISM: Homo sapiens
US-60-455-444-7455

Query Match          50.0%; Score 50; DB 14; Length 328;
Best Local Similarity 55.6%; Pred. No. 7.4;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY      1 VERVDFTNHLDETRRNIN 18
      ||||| |:::|: ||
Db      69 VESVDFAADPESRRKKIN 86

RESULT 10
US-60-465-241-7455
: GENERAL INFORMATION:
: APPLICANT: CARGILL, Michele
: APPLICANT: BEGOVICH, Ann
: TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
: FILE REFERENCE: C1001468
: CURRENT APPLICATION NUMBER: US/60/465,241
: CURRENT FILING DATE: 2003-04-23
: NUMBER OF SEQ ID NOS: 258418
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 7455
: LENGTH: 328
: TYPE: PRT
: ORGANISM: Homo sapiens
US-60-465-241-7455

Query Match          50.0%; Score 50; DB 14; Length 328;
Best Local Similarity 55.6%; Pred. No. 7.4;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY      1 VERVDFTNHLDETRRNIN 18
      ||||| |:::|: ||
Db      69 VESVDFAADPESRRKKIN 86

RESULT 11
PCT-US02-07215B-12
: Sequence 12, Application PC/TUS0207215B
: GENERAL INFORMATION:
: APPLICANT: Immunex Corporation
: APPLICANT: Clarke, Howard RG
: APPLICANT: Dubose, Robert F
: APPLICANT: Wiley, Steven R
: TITLE OF INVENTION: HUMAN SERPIN POLYPEPTIDES
: FILE REFERENCE: 3223-WO
: CURRENT APPLICATION NUMBER: PCT/US02/07215B
: CURRENT FILING DATE: 2002-03-08
: NUMBER OF SEQ ID NOS: 35
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 12
: LENGTH: 390
: TYPE: PRT
: ORGANISM: Homo sapiens
PCT-US02-07215B-12

Query Match          50.0%; Score 50; DB 2; Length 390;
Best Local Similarity 55.6%; Pred. No. 9;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY      1 VERVDFTNHLDETRRNIN 18
      ||||| |:::|: ||
Db      131 VESVDFAADPESRRKKIN 148

```

```
RESULT 12
PCT-US03-17409-175
; Sequence 175, Application PC/TUS0317409
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION
; APPLICANT: RICKERT, Paula K.
; APPLICANT: KRASNOW, Randi
; TITLE OF INVENTION: DIAGNOSTIC MARKERS FOR LUNG CANCER
; FILE REFERENCE: PA-0051 PCT
; CURRENT APPLICATION NUMBER: PCT/US03/17409
; CURRENT FILING DATE: 2003-06-02
; PRIOR FILING DATE: 2002-06-04
; NUMBER OF SEQ ID NOS: 296
; SOFTWARE: PERL Program
; SEQ ID NO 175
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: incyte ID No: 1451328CD1
PCT-US03-17409-175

Query Match          50.0%; Score 50; DB 2; Length 390;
Best Local Similarity 55.6%; Pred. No. 9;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY      1 VERVDFTNHLDETRRNIN 18
      || ||| | |:::| |
Db      131 VESYDFANAPESRKKIN 148

RESULT 13
US-60-448-285-5
; Sequence 5, Application US/60448285
; GENERAL INFORMATION:
; APPLICANT: ASHTON-RICKARDT, PHILIP
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE INHIBITION OF
; TITLE OF INVENTION: CATHESPINS
; FILE REFERENCE: ARCD:390USP1
; CURRENT APPLICATION NUMBER: US/60/448,285
; CURRENT FILING DATE: 2003-02-19
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 5
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-448-285-5

Query Match          50.0%; Score 50; DB 14; Length 390;
Best Local Similarity 55.6%; Pred. No. 9;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY      1 VERVDFTNHLDETRRNIN 18
      || ||| | |:::| |
Db      131 VESYDFANAPESRKKIN 148

RESULT 14
US-09-724-676-57391
; Sequence 57391, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: Patentln version 3.2
; SEQ ID NO 57391
```

```
; LENGTH: 410
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (8)..(8)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-09-724-676-57391

Query Match          50.0%; Score 50; DB 10; Length 410;
Best Local Similarity 55.6%; Pred. No. 9.5;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY      1 VERVDFTNHLDETRRNIN 18
      || ||| | |:::| |
Db      151 VESYDFANAPESRKKIN 168

RESULT 15
US-09-724-676A-57391
; Sequence 57391, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: Patentln version 3.2
; SEQ ID NO 57391
; LENGTH: 410
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (8)..(8)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-09-724-676A-57391

Query Match          50.0%; Score 50; DB 10; Length 410;
Best Local Similarity 55.6%; Pred. No. 9.5;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY      1 VERVDFTNHLDETRRNIN 18
      || ||| | |:::| |
Db      151 VESYDFANAPESRKKIN 168
```

Search completed: July 11, 2003, 12:35:07  
Job time : 19.9545 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:54:32 ; Search time 53.5094 Seconds  
(Without alignments)  
204.833 Million cell updates/sec

Title: US-10-091-442-4

Perfect score: 83  
Sequence: 1 SYIEVEEGTEATATG 17

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending-Patents-AA-Main: \*  
1: /cgn2\_6/ptodata/1/paa/US085\_COMB.pep: \*  
2: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep: \*  
3: /cgn2\_6/ptodata/1/paa/US087\_COMB.pep: \*  
4: /cgn2\_6/ptodata/1/paa/US088\_COMB.pep: \*  
5: /cgn2\_6/ptodata/1/paa/US089\_COMB.pep: \*  
6: /cgn2\_6/ptodata/1/paa/US090\_COMB.pep: \*  
7: /cgn2\_6/ptodata/1/paa/US091\_COMB.pep: \*  
8: /cgn2\_6/ptodata/1/paa/US092\_COMB.pep: \*  
9: /cgn2\_6/ptodata/1/paa/US093\_COMB.pep: \*  
10: /cgn2\_6/ptodata/1/paa/US094\_COMB.pep: \*  
11: /cgn2\_6/ptodata/1/paa/US095\_COMB.pep: \*  
12: /cgn2\_6/ptodata/1/paa/US096\_COMB.pep: \*  
13: /cgn2\_6/ptodata/1/paa/US097\_COMB.pep: \*  
14: /cgn2\_6/ptodata/1/paa/US098\_COMB.pep: \*  
15: /cgn2\_6/ptodata/1/paa/US099\_COMB.pep: \*  
16: /cgn2\_6/ptodata/1/paa/US100\_COMB.pep: \*  
17: /cgn2\_6/ptodata/1/paa/US101\_COMB.pep: \*  
18: /cgn2\_6/ptodata/1/paa/US102\_COMB.pep: \*  
19: /cgn2\_6/ptodata/1/paa/US103\_COMB.pep: \*  
20: /cgn2\_6/ptodata/1/paa/US104\_COMB.pep: \*  
21: /cgn2\_6/ptodata/1/paa/US105\_COMB.pep: \*  
22: /cgn2\_6/ptodata/1/paa/US106\_COMB.pep: \*  
23: /cgn2\_6/ptodata/1/paa/US107\_COMB.pep: \*  
24: /cgn2\_6/ptodata/1/paa/US108\_COMB.pep: \*  
25: /cgn2\_6/ptodata/1/paa/US109\_COMB.pep: \*  
26: /cgn2\_6/ptodata/1/paa/US110\_COMB.pep: \*  
27: /cgn2\_6/ptodata/1/paa/US111\_COMB.pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Length	ID	Description
1	83	100.0	17	US-09-140-719-4
2	83	100.0	17	US-10-091-442-4
3	83	100.0	118	US-60-188-162-4539
4	83	100.0	380	US-08-091-028A-34
5	83	100.0	380	US-09-140-719-34
6	83	100.0	380	US-09-508-997A-2

7	83	100.0	380	US-09-791-537-281	Sequence 281, App
8	83	100.0	380	US-09-936-883A-2	Sequence 2, App1
9	83	100.0	380	US-10-091-442-34	Sequence 34, App1
10	83	100.0	380	US-10-110-030-2	Sequence 2, App1
11	73	88.0	378	US-09-791-537-45859	Sequence 45859, A
12	73	88.0	378	US-10-023-634-84	Sequence 84, App1
13	69	83.1	17	US-08-091-028A-4	Sequence 4, App1
14	69	83.1	17	US-60-196-718-4162	Sequence 4162, Ap
15	69	83.1	102	US-60-196-718-4163	Sequence 4163, Ap
16	69	83.1	116	US-09-834-366-22782	Sequence 22782, A
17	69	83.1	116	US-60-197-873-22782	Sequence 22782, A
18	69	83.1	217	PCT-US01-09369-134	Sequence 134, App
19	69	83.1	217	PCT-US99-01404-8	Sequence 8, App1
20	69	83.1	217	US-09-729-674-134	Sequence 134, App
21	69	83.1	217	US-09-729-674-134	Sequence 169, App
22	69	83.1	217	US-09-729-674-134	Sequence 169, App
23	69	83.1	217	US-09-729-674-134	Sequence 169, App
24	69	83.1	217	US-09-729-674-134	Sequence 169, App
25	69	83.1	217	US-09-729-674-134	Sequence 169, App
26	69	83.1	217	US-09-729-674-134	Sequence 169, App
27	69	83.1	217	US-09-729-674-134	Sequence 169, App
28	69	83.1	217	US-09-729-674-134	Sequence 169, App
29	69	83.1	217	US-09-729-674-134	Sequence 169, App
30	69	83.1	217	US-09-729-674-134	Sequence 169, App
31	69	83.1	217	US-09-729-674-134	Sequence 169, App
32	69	83.1	217	US-09-729-674-134	Sequence 169, App
33	69	83.1	217	US-09-729-674-134	Sequence 169, App
34	69	83.1	217	US-09-729-674-134	Sequence 169, App
35	69	83.1	217	US-09-729-674-134	Sequence 169, App
36	69	83.1	217	US-09-729-674-134	Sequence 169, App
37	69	83.1	217	US-09-729-674-134	Sequence 169, App
38	69	83.1	217	US-09-729-674-134	Sequence 169, App
39	69	83.1	217	US-09-729-674-134	Sequence 169, App
40	69	83.1	217	US-09-729-674-134	Sequence 169, App
41	69	83.1	217	US-09-729-674-134	Sequence 169, App
42	69	83.1	217	US-09-729-674-134	Sequence 169, App
43	69	83.1	217	US-09-729-674-134	Sequence 169, App
44	69	83.1	217	US-09-729-674-134	Sequence 169, App
45	69	83.1	217	US-09-729-674-134	Sequence 169, App

#### ALIGNMENTS

RESULT 1  
US-09-140-719-4  
Sequence 4, Application US/09140719  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSURUOKA, Nobuo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MURA, Kenji  
APPLICANT: ISHIDA, Nobuhiko  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAGUCHI, Kozo  
APPLICANT: YAMAGUCHI, Noriomi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESS: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719

FILING DATE: 08-AUG-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-6620  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 17 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-140-719-4

Query Match 100.0%; Score 83; DB 15; Length 17;  
Best Local Similarity 100.0%; Pred. No. 3.6e-07;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIEVTEEGTEATATG 17  
DB 1 SYIEVTEEGTEATATG 17

## RESULT 2

US-10-091-442-4  
Sequence 4, Application US/10091442

GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi

IMASA, Fuyuki

TSUROOKA, Nobuo

NAKAZATO, Hiroshi

MURA, Kenju

ISHIDA, Nobuhito

KURIHARA, Tatsuya

YAMAUCHI, Kozo

YAMAGUCHI, Nozomi

TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR

NUMBER OF SEQUENCES: 34

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis

STREET: P.O. Box 1404

CITY: Alexandria

STATE: Virginia

COUNTRY: United States

ZIP: 22313-1404

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/091,442

FILING DATE: 07-Mar-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719

FILING DATE: 08-AUG-1998

APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-6620  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 17 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
US-10-091-442-4

Query Match 100.0%; Score 83; DB 24; Length 17;  
Best Local Similarity 100.0%; Pred. No. 3.6e-07;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIEVTEEGTEATATG 17  
DB 1 SYIEVTEEGTEATATG 17

## RESULT 3

US-60-188-162-4539  
Sequence 4539, Application US/60188162

GENERAL INFORMATION:  
APPLICANT: Beasley, Ellen

TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,

TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND

TITLE OF INVENTION: US5S THEREOF

FILE REFERENCE: CLO00518

CURRENT APPLICATION NUMBER: US/60/188,162

CURRENT FILING DATE: 2000-05-09

NUMBER OF SEQ ID NOS: 5094

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 4539

LENGTH: 118

TYPE: PRT

ORGANISM: HUMAN

US-60-188-162-4539

Query Match 100.0%; Score 83; DB 27; Length 118;  
Best Local Similarity 100.0%; Pred. No. 5e-06;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIEVTEEGTEATATG 17  
DB 78 SYIEVTEEGTEATATG 94

## RESULT 4

US-08-091-028A-34  
Sequence 34, Application US/08091028A

GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi

IMASA, Fuyuki

TSUROOKA, Nobuo

NAKAZATO, Hiroshi

MURA, Kenju

ISHIDA, Nobuhito

KURIHARA, Tatsuya

APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, Nozomi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/091,028A  
FILING DATE: 14-JUL-1993  
CLASSIFICATION: 435  
PRIORITY INFORMATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: REA, TERESA STANER  
REGISTRATION NUMBER: 30,427  
REFERENCE/DOCKET NUMBER: 001560-204  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-6620  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-091-028A-34  
Query Match 100.0%; Score 83; DB 4; Length 380;  
Best Local Similarity 100.0%; Pred. No. 2.4e-05;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 SYIEVTEEGTEATATG 17  
DB 325 SYIEVTEEGTEATATG 341  
RESULT 5  
US-09-140-719-34  
Sequence 34, Application US/09140719  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IMASA, Fuyuki  
APPLICANT: TSUROUOKA, Nobuo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, Nobuhiko  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, Nozomi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998  
CLASSIFICATION:  
PRIORITY INFORMATION DATA:  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-2021  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-140-719-34  
Query Match 100.0%; Score 83; DB 15; Length 380;  
Best Local Similarity 100.0%; Pred. No. 2.4e-05;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 SYIEVTEEGTEATATG 17  
DB 325 SYIEVTEEGTEATATG 341  
RESULT 6  
US-09-508-997A-2  
Sequence 2, Application US/09508997A  
GENERAL INFORMATION:  
APPLICANT: MIYATA, Toshio  
TITLE OF INVENTION: Megsin Protein  
FILE REFERENCE: SHIM-004  
CURRENT APPLICATION NUMBER: US/09/508,997A  
CURRENT FILING DATE: 2000-06-27  
PRIOR APPLICATION NUMBER: JP 09-275302  
PRIOR FILING DATE: 1997-09-22  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO: 2  
LENGTH: 380  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-508-997A-2  
Query Match 100.0%; Score 83; DB 19; Length 380;  
Best Local Similarity 100.0%; Pred. No. 2.4e-05;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 SYIEVTEEGTEATATG 17  
DB 325 SYIEVTEEGTEATATG 341  
RESULT 7

US-09-791-537-281  
; Sequence 281, Application US/09791537  
; GENERAL INFORMATION:  
; APPLICANT: Bionomix, Inc.  
; APPLICANT: Deneb, Derek  
; APPLICANT: Danzer, Joseph  
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB  
; FILE REFERENCE: 261/210  
; CURRENT APPLICATION NUMBER: US/09/791,537  
; CURRENT FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 153055  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 281  
; LENGTH: 380  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-791-537-281

Query Match 100.0%; Score 83; DB 21; Length 380;  
Best Local Similarity 100.0%; Pred. No. 2.4e-05;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIEVTEGTEATATG 17  
|||||  
Db 325 SYIEVTEGTEATATG 341

RESULT 8  
US-09-936-883A-2  
; Sequence 2, Application US/09936883A  
; GENERAL INFORMATION:  
; APPLICANT: MIYATA, Toshio  
; TITLE OF INVENTION: A Method for Detecting Megsin Protein and Use  
; FILE REFERENCE: P2-101DIPCT  
; CURRENT APPLICATION NUMBER: US/09/936,883A  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: JP 1999-75305  
; PRIOR FILING DATE: 1999-03-19  
; PRIOR APPLICATION NUMBER: JP 1999-306623  
; PRIOR FILING DATE: 1999-10-28  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 380  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-936-883A-2

Query Match 100.0%; Score 83; DB 23; Length 380;  
Best Local Similarity 100.0%; Pred. No. 2.4e-05;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIEVTEGTEATATG 17  
|||||  
Db 325 SYIEVTEGTEATATG 341

RESULT 9  
US-10-091-442-34  
; Sequence 34, Application US/10091442  
; GENERAL INFORMATION:  
; APPLICANT: TSUJIMOTO, Masafumi  
; APPLICANT: IMASA, Fuyuki  
; APPLICANT: TSUROOKA, Nobuo  
; APPLICANT: NAKAZATO, Hiroshi  
; APPLICANT: MIURA, Kenju  
; APPLICANT: ISHIDA, Nobuhiro  
; APPLICANT: KURIHARA, Tatsuya  
; APPLICANT: YAMAIICHI, Kozo  
; APPLICANT: YAMAGUCHI, Nozomi  
; TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR

NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/091,442  
FILING DATE: 07-Mar-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 34:

US-10-091-442-34

Query Match 100.0%; Score 83; DB 24; Length 380;  
Best Local Similarity 100.0%; Pred. No. 2.4e-05;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYIEVTEGTEATATG 17  
|||||  
Db 325 SYIEVTEGTEATATG 341

RESULT 10  
US-10-110-030-2  
; Sequence 2, Application US/10110030  
; GENERAL INFORMATION:  
; APPLICANT: MIYATA, Toshio  
; TITLE OF INVENTION: ANIMAL MODEL FOR MESENCHYMAL PROLIFERATIVE  
; FILE REFERENCE: SHIM015  
; CURRENT APPLICATION NUMBER: US/10/110,030  
; CURRENT FILING DATE: 2002-04-03  
; PRIOR APPLICATION NUMBER: 11/285736  
; PRIOR FILING DATE: 1999-10-06  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 380  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-110-030-2

Query Match 100.0%; Score 83; DB 25; Length 380;  
Best Local Similarity 100.0%; Pred. No. 2.4e-05;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYEVTGEGTAAATG 17  
DB 325 SYEVTGEGTAAATG 341

## RESULT 11

US-09-791-537-45859  
Sequence 45859, Application US/09791537  
GENERAL INFORMATION:  
APPLICANT: Blonoxix, Inc.  
APPLICANT: Debe, Derek  
TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB  
FILE REFERENCE: 261/210  
CURRENT APPLICATION NUMBER: US/09/791,537  
CURRENT FILING DATE: 2001-02-22  
NUMBER OF SEQ ID NOS: 153055  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 45859  
LENGTH: 378  
TYPE: PR  
ORGANISM: Bos taurus  
US-09-791-537-45859

## Query Match

Best Local Similarity 88.0%; Score 73; DB 21; Length 378;  
Best Local Similarity 82.4%; Pred. No. 0.0013;  
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 SYEVTGEGTAAATG 17  
DB 321 SYEVTGEGTAAATG 337

## RESULT 12

US-10-023-634-84  
Sequence 84, Application US/10023634  
GENERAL INFORMATION:  
APPLICANT: Shinkens, Richard A  
APPLICANT: Coleman, Steven D  
APPLICANT: Spytak, Kimberly A  
APPLICANT: Ballinger, Robert A  
APPLICANT: Guo, XiaoJia  
APPLICANT: Tchernev, Vellizar T  
APPLICANT: Shenoy, Suresh G  
APPLICANT: Li, Li  
APPLICANT: Ellerman, Karen  
APPLICANT: Zerhusen, Bryan D  
APPLICANT: Patlurajan, Meera  
APPLICANT: Casman, Stacie J  
APPLICANT: Boldog, Ferenc  
APPLICANT: Gusev, Vladimir Y  
APPLICANT: Burgess, Catherine E  
APPLICANT: Edinger, Shlomit R  
APPLICANT: Gangoli, Esha A  
APPLICANT: Malyankar, Uriel M  
APPLICANT: Gunther, Erik  
APPLICANT: Smltson, Glenda  
APPLICANT: Millet, Isabelle  
APPLICANT: Gellach, Valerie  
TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of  
FILE REFERENCE: 21402-221  
CURRENT APPLICATION NUMBER: US/10/023,634  
CURRENT FILING DATE: 2002-06-28  
PRIOR APPLICATION NUMBER: 60/256,025  
PRIOR FILING DATE: 2000-12-15  
PRIOR APPLICATION NUMBER: 60/265,163  
PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: 60/272,929  
PRIOR FILING DATE: 2001-03-02  
PRIOR APPLICATION NUMBER: 60/274,864  
PRIOR FILING DATE: 2001-03-09  
PRIOR APPLICATION NUMBER: 60/276,688  
PRIOR FILING DATE: 2001-03-16  
PRIOR APPLICATION NUMBER: 60/277,880  
PRIOR FILING DATE: 2001-03-22  
PRIOR APPLICATION NUMBER: 60/286,409  
PRIOR FILING DATE: 2001-04-25  
PRIOR APPLICATION NUMBER: 60/309,246  
PRIOR FILING DATE: 2001-07-31  
PRIOR APPLICATION NUMBER: 60/315,600  
PRIOR FILING DATE: 2001-08-29  
NUMBER OF SEQ ID NOS: 132  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 84  
LENGTH: 378  
TYPE: PR  
ORGANISM: Bos taurus  
US-10-023-634-84

## Query Match

Best Local Similarity 88.0%; Score 73; DB 24; Length 378;  
Best Local Similarity 82.4%; Pred. No. 0.0013;  
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 SYEVTGEGTAAATG 17  
DB 321 SYEVTGEGTAAATG 337

## RESULT 13

US-08-091-028-4  
Sequence 4, Application US/08091028  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSURUOKA, Nobuo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIDRA, Kenju  
APPLICANT: ISHIDA, Nobuhiko  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, Nozomi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/091,028  
FILING DATE: 14-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Crane-Feury, Sharon E  
REGISTRATION NUMBER: 36,113  
REFERENCE/DOCKET NUMBER: 001560-204  
TELECOMMUNICATION INFORMATION:





;; CURRENT FILING DATE: 2002-04-18  
;; PRIOR APPLICATION NUMBER: US 60/284,770  
;; PRIOR FILING DATE: 2001-04-18  
;; PRIOR APPLICATION NUMBER: US 60/290,492  
;; PRIOR FILING DATE: 2001-05-10  
;; PRIOR APPLICATION NUMBER: US 60/339,245  
;; PRIOR FILING DATE: 2001-11-09  
;; PRIOR APPLICATION NUMBER: US 60/350,666  
;; PRIOR FILING DATE: 2001-11-13  
;; PRIOR APPLICATION NUMBER: US 60/334,370  
;; PRIOR FILING DATE: 2001-11-29  
;; PRIOR APPLICATION NUMBER: US 60/372,246  
;; PRIOR FILING DATE: 2002-04-12  
;; NUMBER OF SEQ ID NOS: 691  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO: 263  
;; LENGTH: 380  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-10-126-052A-263

Query Match  
Best Local Similarity 100.0%; Score 83; DB 12; Length 380;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIEVTEGTEATATG 17  
DB 325 SYIEVTEGTEATATG 341

RESULT 3  
US-60-443-566-3006  
;; Sequence 3006, Application US/60443566  
;; GENERAL INFORMATION:  
;; APPLICANT: CARGILL, Michele  
;; APPLICANT: BEGOVICH, Ann  
;; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
;; FILE REFERENCE: CL001447  
;; CURRENT APPLICATION NUMBER: US/60/443,566  
;; CURRENT FILING DATE: 2003-01-30  
;; NUMBER OF SEQ ID NOS: 25102  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO: 3006  
;; LENGTH: 380  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-60-443-566-3006

Query Match  
Best Local Similarity 100.0%; Score 83; DB 14; Length 380;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIEVTEGTEATATG 17  
DB 325 SYIEVTEGTEATATG 341

RESULT 4  
US-60-455-444-4692  
;; Sequence 4692, Application US/60455444  
;; GENERAL INFORMATION:  
;; APPLICANT: CARGILL, Michele  
;; APPLICANT: BEGOVICH, Ann  
;; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
;; FILE REFERENCE: CL001455  
;; CURRENT APPLICATION NUMBER: US/60/455,444  
;; CURRENT FILING DATE: 2003-03-18  
;; NUMBER OF SEQ ID NOS: 50986  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO: 4692  
;; LENGTH: 380

;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-60-455-444-4692

Query Match  
Best Local Similarity 100.0%; Score 83; DB 14; Length 380;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIEVTEGTEATATG 17  
DB 325 SYIEVTEGTEATATG 341

RESULT 5  
US-60-465-241-4692  
;; Sequence 4692, Application US/60465241  
;; GENERAL INFORMATION:  
;; APPLICANT: CARGILL, Michele  
;; APPLICANT: BEGOVICH, Ann  
;; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
;; FILE REFERENCE: CL001468  
;; CURRENT APPLICATION NUMBER: US/60/465,241  
;; CURRENT FILING DATE: 2003-04-23  
;; NUMBER OF SEQ ID NOS: 258418  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO: 4692  
;; LENGTH: 380  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-60-465-241-4692

Query Match  
Best Local Similarity 100.0%; Score 83; DB 14; Length 380;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIEVTEGTEATATG 17  
DB 325 SYIEVTEGTEATATG 341

RESULT 6  
US-09-936-883C-21  
;; Sequence 21, Application US/09936883C  
;; GENERAL INFORMATION:  
;; APPLICANT: MIYATA, Toshio  
;; TITLE OF INVENTION: A Method for Detecting Megsin protein and Use  
;; FILE REFERENCE: F2-10DIPCT  
;; CURRENT APPLICATION NUMBER: US/09/936,883C  
;; CURRENT FILING DATE: 2001-12-21  
;; PRIOR APPLICATION NUMBER: JP 1999-75305  
;; PRIOR FILING DATE: 1999-03-19  
;; PRIOR APPLICATION NUMBER: JP 1999-306623  
;; PRIOR FILING DATE: 1999-10-28  
;; NUMBER OF SEQ ID NOS: 21  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO: 21  
;; LENGTH: 368  
;; TYPE: PRT  
;; ORGANISM: Mus musculus  
US-09-936-883C-21

Query Match  
Best Local Similarity 83.1%; Score 69; DB 10; Length 368;  
Matches 14; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYIEVTEGTEATATG 16  
DB 313 SYIEVTEGTEATATG 328

RESULT 7



```
US-60-443-566-3717
: Sequence 3717, Application US/60443566
: GENERAL INFORMATION:
: APPLICANT: CARGILL, Michele
: APPLICANT: BEGOVICH, Ann
: TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
: FILE REFERENCE: CLO01447
: CURRENT APPLICATION NUMBER: US/60/443,566
: CURRENT FILING DATE: 2003-01-30
: NUMBER OF SEQ ID NOS: 25102
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 3717
: LENGTH: 372
: TYPE: PRT
: ORGANISM: Homo sapiens
US-60-443-566-3717

Query Match
Best Local Similarity 83.1%; Score 69; DB 14; Length 372;
Matches 13; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 SYIEVTEGTEATATG 17
Db 315 SYLDVSEGTAAATG 331

RESULT 8
US-60-455-444-5630
: Sequence 5630, Application US/60455444
: GENERAL INFORMATION:
: APPLICANT: CARGILL, Michele
: APPLICANT: BEGOVICH, Ann
: TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
: FILE REFERENCE: CLO01455
: CURRENT APPLICATION NUMBER: US/60/455,444
: CURRENT FILING DATE: 2003-03-18
: NUMBER OF SEQ ID NOS: 50986
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 5630
: LENGTH: 372
: TYPE: PRT
: ORGANISM: Homo sapiens
US-60-455-444-5630

Query Match
Best Local Similarity 83.1%; Score 69; DB 14; Length 372;
Matches 13; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 SYIEVTEGTEATATG 17
Db 315 SYLDVSEGTAAATG 331

RESULT 9
US-60-465-241-5630
: Sequence 5630, Application US/60465241
: GENERAL INFORMATION:
: APPLICANT: CARGILL, Michele
: APPLICANT: BEGOVICH, Ann
: TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
: FILE REFERENCE: CLO01468
: CURRENT APPLICATION NUMBER: US/60/465,241
: CURRENT FILING DATE: 2003-04-23
: NUMBER OF SEQ ID NOS: 258418
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 5630
: LENGTH: 372
: TYPE: PRT
: ORGANISM: Homo sapiens
US-60-465-241-5630
```

```
US-10-294-025-946
: Sequence 946, Application US/10294025
: GENERAL INFORMATION:
: APPLICANT: Xu, Jiangchun
: APPLICANT: Stolk, John A.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
: FILE REFERENCE: 210121.427C29
: CURRENT APPLICATION NUMBER: US/10/294,025
: CURRENT FILING DATE: 2002-11-12
: NUMBER OF SEQ ID NOS: 1038
: SOFTWARE: FASTSEQ for Windows Version 3.0
: SEQ ID NO 946
: LENGTH: 379
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-294-025-946

Query Match
Best Local Similarity 83.1%; Score 69; DB 12; Length 379;
Matches 13; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 SYIEVTEGTEATATG 17
Db 322 SYLDVSEGTAAATG 338

RESULT 10
US-10-294-025-946
: Sequence 946, Application US/10294025
: GENERAL INFORMATION:
: APPLICANT: Xu, Jiangchun
: APPLICANT: Stolk, John A.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
: FILE REFERENCE: 210121.427C29
: CURRENT APPLICATION NUMBER: US/10/294,025
: CURRENT FILING DATE: 2002-11-12
: NUMBER OF SEQ ID NOS: 1038
: SOFTWARE: FASTSEQ for Windows Version 3.0
: SEQ ID NO 946
: LENGTH: 379
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-294-025-946

Query Match
Best Local Similarity 83.1%; Score 69; DB 12; Length 392;
Matches 13; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 SYIEVTEGTEATATG 17
Db 335 SYLDVSEGTAAATG 351

RESULT 11
US-10-198-070-62
: Sequence 62, Application US/10198070
: GENERAL INFORMATION:
: APPLICANT: AVERBACK, PAUL
: APPLICANT: GEMMELL, JACK
: TITLE OF INVENTION: PEPTIDES EFFECTIVE IN THE TREATMENT OF TUMORS AND OTHER
: TITLE OF INVENTION: CONDITIONS REQUIRING THE REMOVAL OR DESTRUCTION OF
: FILE REFERENCE: 59003.000008
: CURRENT APPLICATION NUMBER: US/10/198,070
: CURRENT FILING DATE: 2002-07-19
: PRIOR APPLICATION NUMBER: 60/306,161
: PRIOR FILING DATE: 2001-07-19
: PRIOR APPLICATION NUMBER: 60/306,150
: PRIOR FILING DATE: 2001-07-19
: PRIOR APPLICATION NUMBER: 60/331,477
: PRIOR FILING DATE: 2001-11-16
: NUMBER OF SEQ ID NOS: 125
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 62
: LENGTH: 392
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-198-070-62

Query Match
Best Local Similarity 83.1%; Score 69; DB 12; Length 392;
Matches 13; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 SYIEVTEGTEATATG 17
Db 335 SYLDVSEGTAAATG 351
```

RESULT 12  
US-10-198-070-67  
; Sequence 67, Application US/10198070  
; GENERAL INFORMATION:  
; APPLICANT: AVERBACK, PAUL  
; APPLICANT: GEMMELL, JACK  
; TITLE OF INVENTION: PEPTIDES EFFECTIVE IN THE TREATMENT OF TUMORS AND OTHER  
; TITLE OF INVENTION: CONDITIONS REQUIRING THE REMOVAL OR DESTRUCTION OF  
; TITLE OF INVENTION: CELLS  
; FILE REFERENCE: 59003.000008  
; CURRENT APPLICATION NUMBER: US/10/198,070  
; CURRENT FILING DATE: 2002-07-19  
; PRIOR APPLICATION NUMBER: 60/306,161  
; PRIOR FILING DATE: 2001-07-19  
; PRIOR APPLICATION NUMBER: 60/306,150  
; PRIOR FILING DATE: 2001-07-19  
; PRIOR APPLICATION NUMBER: 60/331,477  
; PRIOR FILING DATE: 2001-11-16  
; NUMBER OF SEQ ID NOS: 125  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 67  
; LENGTH: 392  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-198-070-67

Query Match 83.1%; Score 69; DB 12; Length 392;  
Best Local Similarity 76.5%; Pred. No. 0.0081;  
Matches 13; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SYIEVTEGTEATATG 17  
Db 335 SYLDVSEEGTEAAATG 351  
RESULT 13  
US-10-198-070-70  
; Sequence 70, Application US/10198070  
; GENERAL INFORMATION:  
; APPLICANT: AVERBACK, PAUL  
; APPLICANT: GEMMELL, JACK  
; TITLE OF INVENTION: PEPTIDES EFFECTIVE IN THE TREATMENT OF TUMORS AND OTHER  
; TITLE OF INVENTION: CONDITIONS REQUIRING THE REMOVAL OR DESTRUCTION OF  
; TITLE OF INVENTION: CELLS  
; FILE REFERENCE: 59003.000008  
; CURRENT APPLICATION NUMBER: US/10/198,070  
; CURRENT FILING DATE: 2002-07-19  
; PRIOR APPLICATION NUMBER: 60/306,161  
; PRIOR FILING DATE: 2001-07-19  
; PRIOR APPLICATION NUMBER: 60/306,150  
; PRIOR FILING DATE: 2001-07-19  
; PRIOR APPLICATION NUMBER: 60/331,477  
; PRIOR FILING DATE: 2001-11-16  
; NUMBER OF SEQ ID NOS: 125  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 70  
; LENGTH: 392  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-198-070-70

Query Match 83.1%; Score 69; DB 12; Length 392;  
Best Local Similarity 76.5%; Pred. No. 0.0081;  
Matches 13; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SYIEVTEGTEATATG 17  
Db 335 SYLDVSEEGTEAAATG 351

RESULT 14  
PCT-US02-07215B-14

; Sequence 14, Application PC/TUS0207215B  
; GENERAL INFORMATION:  
; APPLICANT: Immunex Corporation  
; APPLICANT: Clarke, Howard RG  
; APPLICANT: Dubose, Robert F  
; APPLICANT: Wiley, Steven R  
; TITLE OF INVENTION: HUMAN SERPIN POLYPEPTIDES  
; FILE REFERENCE: 3223-WO  
; CURRENT APPLICATION NUMBER: PCT/US02/07215B  
; CURRENT FILING DATE: 2002-03-08  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 14  
; LENGTH: 617  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (90)..(90)  
; OTHER INFORMATION: unsure  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (215)..(215)  
; OTHER INFORMATION: unsure  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (225)..(225)  
; OTHER INFORMATION: unsure  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (232)..(233)  
; OTHER INFORMATION: unsure  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (289)..(289)  
; OTHER INFORMATION: unsure  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (309)..(309)  
; OTHER INFORMATION: unsure  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (318)..(318)  
; OTHER INFORMATION: unsure  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (322)..(322)  
; OTHER INFORMATION: unsure  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (331)..(332)  
; OTHER INFORMATION: unsure  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (367)..(367)  
; OTHER INFORMATION: unsure  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (369)..(369)  
; OTHER INFORMATION: unsure  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (374)..(374)  
; OTHER INFORMATION: unsure  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (378)..(378)  
; OTHER INFORMATION: unsure  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (425)..(425)  
; OTHER INFORMATION: unsure  
; FEATURE:

Fri Jul 11 15:11:00 2003

NAME/KEY: MISC\_FEATURE  
LOCATION: (430)..(430)  
OTHER INFORMATION: unsure  
PCT-US02-07215B-14

Query Match 83.1%; Score 69; DB 2; Length 617;  
Best Local Similarity 76.5%; Pred. No. 0.013;  
Matches 13; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 SYIEVEEGTEATAATG 17  
||:::|||||  
Db 560 SYLDVSEEGTEAAATG 576

RESULT 15  
US-10-294-025-947  
Sequence 947, Application US/10294025  
GENERAL INFORMATION:  
APPLICANT: Xu, Jiangchun  
APPLICANT: Stolk, John A.  
APPLICANT: Kajos, Michael D.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER  
FILE REFERENCE: 210121.427C29  
CURRENT APPLICATION NUMBER: US/10/294,025  
CURRENT FILING DATE: 2002-11-12  
NUMBER OF SEQ ID NOS: 1038  
SOFTWARE: FASTSEQ for Windows Version 3.0  
SEQ ID NO 947  
LENGTH: 617  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: 90, 215, 225, 232, 233, 289, 309, 318, 322, 331, 332, 367,  
LOCATION: 369, 374, 378, 425, 430  
OTHER INFORMATION: Xaa - Any Amino Acid  
US-10-294-025-947

Query Match 83.1%; Score 69; DB 12; Length 617;  
Best Local Similarity 76.5%; Pred. No. 0.013;  
Matches 13; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 SYIEVEEGTEATAATG 17  
||:::|||||  
Db 560 SYLDVSEEGTEAAATG 576

Search completed: July 11, 2003, 12:35:07  
Job time : 16.0751 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: July 11, 2003, 11:54:32 : Search time 28.3285 Seconds  
(without alignments)  
204.833 Million cell updates/sec

Title: US-10-091-442-5

Perfect score: 44

Sequence: 1 QYLRALGLK 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

Pending\_Patents\_AA\_Main:\*  
1: /cgn2\_6/ptodata/1/paa/PCTUS.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/paa/US06.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/paa/US07.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/paa/US08.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/paa/US09.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/paa/US082.COMB.pep:\*  
7: /cgn2\_6/ptodata/1/paa/US083.COMB.pep:\*  
8: /cgn2\_6/ptodata/1/paa/US084.COMB.pep:\*  
9: /cgn2\_6/ptodata/1/paa/US085.COMB.pep:\*  
10: /cgn2\_6/ptodata/1/paa/US086.COMB.pep:\*  
11: /cgn2\_6/ptodata/1/paa/US087.COMB.pep:\*  
12: /cgn2\_6/ptodata/1/paa/US088.COMB.pep:\*  
13: /cgn2\_6/ptodata/1/paa/US089.COMB.pep:\*  
14: /cgn2\_6/ptodata/1/paa/US090.COMB.pep:\*  
15: /cgn2\_6/ptodata/1/paa/US091.COMB.pep:\*  
16: /cgn2\_6/ptodata/1/paa/US092.COMB.pep:\*  
17: /cgn2\_6/ptodata/1/paa/US093.COMB.pep:\*  
18: /cgn2\_6/ptodata/1/paa/US094.COMB.pep:\*  
19: /cgn2\_6/ptodata/1/paa/US095.COMB.pep:\*  
20: /cgn2\_6/ptodata/1/paa/US096.COMB.pep:\*  
21: /cgn2\_6/ptodata/1/paa/US097.COMB.pep:\*  
22: /cgn2\_6/ptodata/1/paa/US098.COMB.pep:\*  
23: /cgn2\_6/ptodata/1/paa/US099.COMB.pep:\*  
24: /cgn2\_6/ptodata/1/paa/US100.COMB.pep:\*  
25: /cgn2\_6/ptodata/1/paa/US101.COMB.pep:\*  
26: /cgn2\_6/ptodata/1/paa/US102.COMB.pep:\*  
27: /cgn2\_6/ptodata/1/paa/US60.COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44	100.0	9	US-09-140-719-5	Sequence 5, Appl
2	44	100.0	9	US-10-091-442-5	Sequence 5, Appl
3	44	100.0	78	US-60-160-203-4173	Sequence 4173, Ap
4	44	100.0	118	US-60-188-162-4539	Sequence 4539, Ap
5	44	100.0	380	US-08-091-028A-34	Sequence 34, Appl
6	44	100.0	380	US-09-140-719-34	Sequence 34, Appl

7	44	100.0	380	US-09-508-997A-2	Sequence 2, Appl
8	44	100.0	380	US-09-791-537-281	Sequence 281, App
9	44	100.0	380	US-09-936-883A-2	Sequence 2, Appl
10	44	100.0	380	US-10-091-442-34	Sequence 34, Appl
11	44	100.0	380	US-10-110-030-2	Sequence 2, Appl
12	39	88.6	9	US-08-091-028A-5	Sequence 5, Appl
13	39	88.6	9	US-08-091-028A-5	Sequence 5, Appl
14	35	79.5	598	US-09-760-469-1213	Sequence 1213, Ap
15	35	79.5	598	US-10-216-583-1213	Sequence 1213, Ap
16	35	79.5	694	US-09-856-679-12	Sequence 12, Appl
17	35	79.5	694	US-09-856-679-12	Sequence 12, Appl
18	34	77.3	80	US-09-791-537-48914	Sequence 5, Appl
19	34	77.3	210	US-09-791-537-92802	Sequence 48914, A
20	34	77.3	213	PCT-US01-08631-36588	Sequence 92802, A
21	34	77.3	270	US-09-417-507-34770	Sequence 36588, A
22	34	77.3	313	US-09-248-796-17065	Sequence 17065, A
23	34	77.3	313	US-09-248-796-17065	Sequence 17065, A
24	34	77.3	337	US-60-096-409-17065	Sequence 8700, Ap
25	34	77.3	350	US-09-791-537-8066	Sequence 8066, Ap
26	34	77.3	350	US-09-791-537-9963	Sequence 9963, Ap
27	34	77.3	350	US-09-847-208-53	Sequence 53, Appl
28	34	77.3	350	US-09-847-208-53	Sequence 53, Appl
29	34	77.3	409	US-10-179-131-7870	Sequence 7870, Ap
30	34	77.3	434	US-10-179-131-5552	Sequence 5552, Ap
31	34	77.3	597	US-09-540-209B-9878	Sequence 9878, Ap
32	33	75.0	250	US-09-184-082-2	Sequence 2, Appl
33	33	75.0	537	US-60-360-039-20450	Sequence 20450, A
34	33	75.0	1366	PCT-US98-00615-22	Sequence 22, Appl
35	33	75.0	1366	US-08-781-734-22	Sequence 22, Appl
36	33	75.0	1366	US-08-781-734-22	Sequence 22, Appl
37	33	75.0	1465	US-10-083-357-1310	Sequence 1310, Ap
38	33	75.0	1465	US-60-333-726-1310	Sequence 1310, Ap
39	33	75.0	1890	PCT-US98-00615-88	Sequence 88, Appl
40	32	72.7	80	US-09-791-537-14885	Sequence 14885, Ap
41	32	72.7	92	US-60-195-052-1080	Sequence 1080, Ap
42	32	72.7	92	US-60-195-052-1149	Sequence 1149, Ap
43	32	72.7	96	US-60-195-052-1000	Sequence 1000, Ap
44	32	72.7	97	US-60-213-178-600	Sequence 600, App
45	32	72.7	102	US-60-202-183-189	Sequence 189, App

## ALIGNMENTS

RESULT 1  
US-09-140-719-5  
; Sequence 5, Application US/09140719  
; GENERAL INFORMATION:  
; APPLICANT: TSUJIMOTO, Masaaki  
; APPLICANT: IMASA, Fuyuki  
; APPLICANT: TSUBOHOKA, Nobuo  
; APPLICANT: NAKAZATO, Hiroshi  
; APPLICANT: MIDURA, Kenju  
; APPLICANT: ISHIDA, Nobuhiko  
; APPLICANT: KURIHARA, Tatsuya  
; APPLICANT: YAMAGUCHI, Kozo  
; APPLICANT: YAMAGUCHI, Nozomi  
; TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
; NUMBER OF SEQUENCES: 34  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: P.O. Box 1404  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/140,719

FILED DATE: 08-AUG-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: MCGOWAN, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-6620  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-140-719-5

Query Match 100.0%; Score 44; DB 15; Length 9;  
Best Local Similarity 100.0%; Pred. No. 4.2e+06;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYLRALGLK 9  
DB 1 QYLRALGLK 9

RESULT 2  
US-10-091-442-5  
Sequence 5, Application US/10091442  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
IMASA, Fuyuki  
TSURUOKA, Nobuo  
NAKAZATO, Hiroshi  
MIURA, Kenju  
ISHIDA, Nobuhiko  
KURIHARA, Tatsuya  
YAMAICHI, Kozo  
YAMAGUCHI, Nozomi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/091,442  
FILING DATE: 07-Mar-2002  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998

APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: MCGOWAN, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-6620  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
US-10-091-442-5

Query Match 100.0%; Score 44; DB 24; Length 9;  
Best Local Similarity 100.0%; Pred. No. 4.2e+06;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYLRALGLK 9  
DB 1 QYLRALGLK 9

RESULT 3  
US-60-160-203-4173  
Sequence 4173, Application US/60160203  
GENERAL INFORMATION:  
APPLICANT: BONAZZI, VIVIAN  
TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS AND  
FILE REFERENCE: CL000116  
CURRENT APPLICATION NUMBER: US/60/160,203  
CURRENT FILING DATE: 1999-10-19  
NUMBER OF SEQ ID NOS: 6374  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 4173  
LENGTH: 78  
TYPE: PRT  
ORGANISM: HUMAN  
US-60-160-203-4173

Query Match 100.0%; Score 44; DB 27; Length 78;  
Best Local Similarity 100.0%; Pred. No. 1.1;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYLRALGLK 9  
DB 42 QYLRALGLK 50

RESULT 4  
US-60-188-162-4539  
Sequence 4539, Application US/60188162  
GENERAL INFORMATION:  
APPLICANT: Beasley, Ellen  
TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND  
FILE REFERENCE: CL000518  
CURRENT APPLICATION NUMBER: US/60/188,162  
CURRENT FILING DATE: 2000-05-09

NUMBER OF SEQ ID NOS: 5094  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 4539  
LENGTH: 118  
TYPE: PRT  
ORGANISM: HUMAN  
US-60-188-162-4539

Query Match 100.0%; Score 44; DB 27; Length 118;  
Best Local Similarity 100.0%; Pred. No. 1.7;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 OYLRALGLK 9  
|||||  
DB 42 OYLRALGLK 50

RESULT 5  
US-08-091-028A-34

Sequence 34, Application US/08091028A  
GENERAL INFORMATION:

APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSURUOKA, Nobuo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, Nobuhito  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAGUCHI, Kozo  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/091,028A  
FILING DATE: 14-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: REA, TERESA STANEK  
REGISTRATION NUMBER: 30,427  
REFERENCE/DOCKET NUMBER: 001560-204  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-6620  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-091-028A-34

Query Match 100.0%; Score 44; DB 4; Length 380;  
Best Local Similarity 100.0%; Pred. No. 6;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 OYLRALGLK 9  
|||||  
DB 289 OYLRALGLK 297

RESULT 6  
US-09-140-719-34

Sequence 34, Application US/09140719  
GENERAL INFORMATION:

APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSURUOKA, Nobuo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, Nobuhito  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAGUCHI, Kozo  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021

INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-140-719-34

Query Match 100.0%; Score 44; DB 15; Length 380;  
Best Local Similarity 100.0%; Pred. No. 6;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 OYLRALGLK 9  
|||||  
DB 289 OYLRALGLK 297

RESULT 7

US-09-508-997A-2  
; Sequence 2, Application US/09508997A  
; GENERAL INFORMATION:  
; APPLICANT: MIYATA, Toshio  
; TITLE OF INVENTION: Megsin Protein  
; FILE REFERENCE: SHIM-004  
; CURRENT APPLICATION NUMBER: US/09/508,997A  
; CURRENT FILING DATE: 2000-06-27  
; PRIOR APPLICATION NUMBER: JP 09-275302  
; PRIOR FILING DATE: 1997-09-22  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 380  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-508-997A-2

Query Match  
Best Local Similarity 100.0%; Score 44; DB 19; Length 380;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 QYLRALGLK 9  
DB 289 QYLRALGLK 297

RESULT 8  
US-09-791-537-281  
; Sequence 281, Application US/09791537  
; GENERAL INFORMATION:  
; APPLICANT: Biomomix, Inc.  
; APPLICANT: Debe, Derek  
; APPLICANT: Danzer, Joseph  
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS  
; FILE REFERENCE: 261/210  
; CURRENT APPLICATION NUMBER: US/09/791,537  
; CURRENT FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 153055  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 281  
; LENGTH: 380  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-791-537-281

Query Match  
Best Local Similarity 100.0%; Score 44; DB 21; Length 380;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 QYLRALGLK 9  
DB 289 QYLRALGLK 297

RESULT 9  
US-09-936-883A-2  
; Sequence 2, Application US/09936883A  
; GENERAL INFORMATION:  
; APPLICANT: MIYATA, Toshio  
; TITLE OF INVENTION: A Method for Detecting Megsin Protein and Use  
; FILE REFERENCE: F2-101DIPACT  
; CURRENT APPLICATION NUMBER: US/09/936,883A  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: JP 1999-75305  
; PRIOR FILING DATE: 1999-03-19  
; PRIOR APPLICATION NUMBER: JP 1999-306623  
; PRIOR FILING DATE: 1999-10-28  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2

LENGTH: 380  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-936-883A-2

Query Match  
Best Local Similarity 100.0%; Score 44; DB 23; Length 380;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 QYLRALGLK 9  
DB 289 QYLRALGLK 297

RESULT 10  
US-10-091-442-34  
; Sequence 34, Application US/10091442  
; GENERAL INFORMATION:  
; APPLICANT: TSUTSUMOTO, Masafumi  
; APPLICANT: IWASA, Fuyuki  
; APPLICANT: TSUROOKA, Nobuo  
; APPLICANT: NAKAZATO, Hiroshi  
; APPLICANT: MURA, Kenju  
; APPLICANT: ISHIDA, Nobuhito  
; APPLICANT: KURIHARA, Tatsuya  
; APPLICANT: YAMACHI, Kozo  
; APPLICANT: YAMAGUCHI, Nozomi  
; TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
; NUMBER OF SEQUENCES: 34  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: P.O. Box 1404  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/091,442  
; FILING DATE: 07-Mar-2002  
; CLASSIFICATION: <unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/09/140,719  
; FILING DATE: 08-AUG-1998  
; APPLICATION NUMBER: US 08/474,661  
; FILING DATE: 07-JUN-1995  
; APPLICATION NUMBER: US 08/091,028  
; FILING DATE: 14-JUL-1993  
; APPLICATION NUMBER: JP 4-212305  
; FILING DATE: 17-JUL-1992  
; APPLICATION NUMBER: JP 6-067339  
; FILING DATE: 04-MAR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McGowan, Malcolm K.  
; REGISTRATION NUMBER: 39,300  
; REFERENCE/DOCKET NUMBER: 001560-247  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 34:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 380 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 34:  
US-10-091-442-34

Query Match  
100.0%; Score 44; DB 24; Length 380;



Best Local Similarity 100.0%; Pred. No. 6;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 OYLRALGLK 9  
Db 289 OYLRALGLK 297

RESULT 11  
US-10-110-030-2

Sequence 2, Application US/10110030  
GENERAL INFORMATION:  
APPLICANT: Miyata, Toshio  
TITLE OF INVENTION: ANIMAL MODEL FOR MESENTERIAL PROLIFERATIVE  
FILE REFERENCE: SHIM015  
CURRENT APPLICATION NUMBER: US/10/110-030  
CURRENT FILING DATE: 2002-04-03  
PRIOR APPLICATION NUMBER: 11/285736  
PRIOR FILING DATE: 1999-10-06  
NUMBER OF SEQ ID NOS: 15  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 380  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-110-030-2

Query Match 100.0%; Score 44; DB 25; Length 380;  
Best Local Similarity 100.0%; Pred. No. 6;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 OYLRALGLK 9  
Db 289 OYLRALGLK 297

RESULT 12  
US-08-091-028-5

Sequence 5, Application US/08091028  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IMASA, Fuyuki  
APPLICANT: TSURUOKA, Nobuo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MURA, Kenju  
APPLICANT: ISHIDA, Nobuhiko  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, Nozomi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/091.028  
FILING DATE: 14-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-067339

FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Crane-Feuyl, Sharon E  
REGISTRATION NUMBER: 36,113  
REFERENCE/DOCKET NUMBER: 001560-204  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-6620  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-091-028-5

OY 2 YLRALGLK 9  
Db 2 YLRALGLK 9

RESULT 13  
US-08-091-028A-5

Sequence 5, Application US/08091028A  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IMASA, Fuyuki  
APPLICANT: TSURUOKA, Nobuo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MURA, Kenju  
APPLICANT: ISHIDA, Nobuhiko  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, Nozomi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/091.028A  
FILING DATE: 14-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: REA, TERESA STANER  
REGISTRATION NUMBER: 30,427  
REFERENCE/DOCKET NUMBER: 001560-204  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-6620  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids

TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-091-028A-5

Query Match  
Best Local Similarity 88.6%; Score 39; DB 4; Length 9;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 YLRALGLK 9  
Db 2 YLRALGLK 9

RESULT 14  
US-09-760-469-1213  
Sequence 1213, Application US/09760469  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PT254  
CURRENT APPLICATION NUMBER: US/09/760,469  
CURRENT FILING DATE: 2001-01-16  
Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 1983  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1213  
LENGTH: 598  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-760-469-1213

Query Match  
Best Local Similarity 79.5%; Score 35; DB 21; Length 598;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYLRALGL 8  
Db 455 YLRALGL 462

RESULT 15  
US-10-216-583-1213  
Sequence 1213, Application US/10216583  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PT254CIN  
CURRENT APPLICATION NUMBER: US/10/216,583  
CURRENT FILING DATE: 2002-08-12  
PRIOR APPLICATION NUMBER: 09/760,469  
PRIOR FILING DATE: 2001-01-16  
PRIOR APPLICATION NUMBER: 60/179,065  
PRIOR FILING DATE: 2000-01-31  
PRIOR APPLICATION NUMBER: 60/180,628  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: 60/214,886  
PRIOR FILING DATE: 2000-06-28  
PRIOR APPLICATION NUMBER: 60/217,487  
PRIOR FILING DATE: 2000-07-11  
PRIOR APPLICATION NUMBER: 60/225,758  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/220,963  
PRIOR FILING DATE: 2000-07-26  
PRIOR APPLICATION NUMBER: 60/217,496  
PRIOR FILING DATE: 2000-07-11  
PRIOR APPLICATION NUMBER: 60/225,447  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/218,290  
PRIOR FILING DATE: 2000-07-14  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 1983

SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1213  
LENGTH: 598  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-216-583-1213

Query Match  
Best Local Similarity 79.5%; Score 35; DB 26; Length 598;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 QYLRALGL 8  
Db 455 YLRALGL 462

Search completed: July 11, 2003, 12:27:33  
Job time : 30.3285 secs

```

: APPLICANT: Murray, Richard
: APPLICANT: Eos Biotechnology, Inc.
: TITLE OF INVENTION: Methods of Diagnosis of Lung Cancer, Compositions and
: TITLE OF INVENTION: Methods of Screening for Modulators of Lung Cancer
: FILE REFERENCE: 018501-001530US
: CURRENT APPLICATION NUMBER: US/10/126,052A

```

;; CURRENT FILING DATE: 2002-04-18  
;; PRIOR APPLICATION NUMBER: US 60/284,770  
;; PRIOR FILING DATE: 2001-04-18  
;; PRIOR APPLICATION NUMBER: US 60/290,492  
;; PRIOR FILING DATE: 2001-05-10  
;; PRIOR APPLICATION NUMBER: US 60/339,245  
;; PRIOR FILING DATE: 2001-11-09  
;; PRIOR APPLICATION NUMBER: US 60/350,666  
;; PRIOR FILING DATE: 2001-11-13  
;; PRIOR APPLICATION NUMBER: US 60/334,370  
;; PRIOR FILING DATE: 2001-11-29  
;; PRIOR APPLICATION NUMBER: US 60/372,246  
;; PRIOR FILING DATE: 2002-04-12  
;; NUMBER OF SEQ ID NOS: 691  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 263  
;; LENGTH: 380  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-10-126-052A-263

Query Match 100.0%; Score 44; DB 12; Length 380;  
Best Local Similarity 100.0%; Pred. No. 1;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 OYLRALGLK 9  
Db 289 OYLRALGLK 297

RESULT 3  
US-60-443-566-3006  
;; Sequence 3006, Application US/60443566  
;; GENERAL INFORMATION:  
;; APPLICANT: CARGILL, Michele  
;; APPLICANT: BEGOVICH, Ann  
;; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
;; FILE REFERENCE: CL001447  
;; CURRENT APPLICATION NUMBER: US/60/443,566  
;; CURRENT FILING DATE: 2003-01-30  
;; NUMBER OF SEQ ID NOS: 25102  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 3006  
;; LENGTH: 380  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-60-443-566-3006

Query Match 100.0%; Score 44; DB 14; Length 380;  
Best Local Similarity 100.0%; Pred. No. 1;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 OYLRALGLK 9  
Db 289 OYLRALGLK 297

RESULT 4  
US-60-455-444-4692  
;; Sequence 4692, Application US/60455444  
;; GENERAL INFORMATION:  
;; APPLICANT: CARGILL, Michele  
;; APPLICANT: BEGOVICH, Ann  
;; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
;; FILE REFERENCE: CL001455  
;; CURRENT APPLICATION NUMBER: US/60/455,444  
;; CURRENT FILING DATE: 2003-03-18  
;; NUMBER OF SEQ ID NOS: 50986  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 4692  
;; LENGTH: 380

;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-60-455-444-4692

Query Match 100.0%; Score 44; DB 14; Length 380;  
Best Local Similarity 100.0%; Pred. No. 1;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 OYLRALGLK 9  
Db 289 OYLRALGLK 297

RESULT 5  
US-60-465-241-4692  
;; Sequence 4692, Application US/60465241  
;; GENERAL INFORMATION:  
;; APPLICANT: CARGILL, Michele  
;; APPLICANT: BEGOVICH, Ann  
;; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
;; FILE REFERENCE: CL001468  
;; CURRENT APPLICATION NUMBER: US/60/465,241  
;; CURRENT FILING DATE: 2003-04-23  
;; NUMBER OF SEQ ID NOS: 258418  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 4692  
;; LENGTH: 380  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-60-465-241-4692

Query Match 100.0%; Score 44; DB 14; Length 380;  
Best Local Similarity 100.0%; Pred. No. 1;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 OYLRALGLK 9  
Db 289 OYLRALGLK 297

RESULT 6  
US-10-437-963-150051  
;; Sequence 150051, Application US/10437963  
;; GENERAL INFORMATION:  
;; APPLICANT: La Rosa, Thomas J.  
;; APPLICANT: Kovalic, David K.  
;; APPLICANT: Zhou, Yihua  
;; APPLICANT: Cao, Yongwei  
;; APPLICANT: Wu, Wei  
;; APPLICANT: Boukharov, Andrey A.  
;; APPLICANT: Barbazuk, Brad  
;; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
;; FILE REFERENCE: 38-21(53221)B  
;; CURRENT APPLICATION NUMBER: US/10/437,963  
;; CURRENT FILING DATE: 2003-05-14  
;; NUMBER OF SEQ ID NOS: 204966  
;; SEQ ID NO 150051  
;; LENGTH: 730  
;; TYPE: PRT  
;; ORGANISM: Oryza sativa  
;; FEATURE:  
;; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_50323C.1.pep  
US-10-437-963-150051

Query Match 81.8%; Score 36; DB 12; Length 730;  
Best Local Similarity 77.8%; Pred. No. 96;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 OYLRALGLK 9  
Db 289 OYLRALGLK 297

Db 524 QYLRAIGOK 532

## RESULT 7

US-10-437-963-117663

; Sequence 117663, Application US/10437963

; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; APPLICANT: Wu, Wei

; APPLICANT: Boukharov, Andrey A.

; APPLICANT: Barbazuk, Brad

; APPLICANT: Li, Ping

; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with

; FILE REFERENCE: 38-21(5322)B

; CURRENT APPLICATION NUMBER: US/10/437,963

; CURRENT FILING DATE: 2003-05-14

; NUMBER OF SEQ ID NOS: 204966

; SEQ ID NO 117663

; LENGTH: 1293

; TYPE: PRT

; ORGANISM: Oryza sativa

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_21048C.1.pep

US-10-437-963-117663

## Query Match

Best Local Similarity 79.5%; Score 35; DB 12; Length 1293;

Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QYLRAIGLK 9

Db 593 RYLKALGIR 601

## RESULT 8

US-10-369-493-8700

; Sequence 8700, Application US/10369493

; GENERAL INFORMATION:

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 8700

; LENGTH: 337

; TYPE: PRT

; ORGANISM: Ralstonia metallidurans

US-10-369-493-8700

## Query Match

Best Local Similarity 77.3%; Score 34; DB 12; Length 337;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QYLRAIGLK 9

Db 182 QYARAGMLR 190

## RESULT 9

US-10-282-122A-47577

; Sequence 47577, Application US/10282122A

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John

; APPLICANT: Carr, Grant

; APPLICANT: Yamamoto, Robert

; APPLICANT: Forsyth, R.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELTRA 034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See file Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 47577

; LENGTH: 282

; TYPE: PRT

; ORGANISM: Burkholderia cepacia

US-10-282-122A-47577

## Query Match

Best Local Similarity 75.0%; Score 33; DB 12; Length 282;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QYLRAIGLK 9

Db 38 QYLKSTGLR 46

## RESULT 10

US-10-369-493-20450

; Sequence 20450, Application US/10369493

; GENERAL INFORMATION:

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 20450

LENGTH: 537  
TYPE: PRT  
ORGANISM: Rhodospseudomonas palustris  
US-10-369-493-20450

Query Match 75.0%; Score 33; DB 12; Length 537;  
Best Local Similarity 87.5%; Pred. No. 2.9e+02;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LRALGLK 9  
DB 81 YLRAGLX 88

RESULT 11  
US-10-437-963-115178  
; Sequence 115178, Application US/10437963  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovall, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Bouharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437, 963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 115178  
; LENGTH: 582  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_187C.1.pep  
US-10-437-963-115178

Query Match 75.0%; Score 33; DB 12; Length 582;  
Best Local Similarity 66.7%; Pred. No. 3.1e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 QYLRAGLX 9  
DB 356 QYARVGLK 364

RESULT 12  
US-10-425-114-60805  
; Sequence 60805, Application US/10425114  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jindong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovall, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(5313)B  
; CURRENT APPLICATION NUMBER: US/10/425, 114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 60805  
; LENGTH: 124  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3356-048-F4\_FLI.pep  
US-10-425-114-60805

Query Match 72.7%; Score 32; DB 12; Length 124;  
Best Local Similarity 100.0%; Pred. No. 88;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LRALGLK 9  
DB 111 LRALGLK 117

RESULT 13  
US-10-425-114-59319  
; Sequence 59319, Application US/10425114  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jindong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovall, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(5313)B  
; CURRENT APPLICATION NUMBER: US/10/425, 114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 59319  
; LENGTH: 127  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB189-027-D3\_FLI.pep  
US-10-425-114-59319

Query Match 72.7%; Score 32; DB 12; Length 127;  
Best Local Similarity 100.0%; Pred. No. 90;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LRALGLK 9  
DB 114 LRALGLK 120

RESULT 14  
US-10-425-114-60833  
; Sequence 60833, Application US/10425114  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jindong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovall, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(5313)B  
; CURRENT APPLICATION NUMBER: US/10/425, 114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 60833  
; LENGTH: 150  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3354-105-B8\_FLI.pep  
US-10-425-114-60833

Query Match 72.7%; Score 32; DB 12; Length 150;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LRALGLK 9  
DB 137 LRALGLK 143

RESULT 15

US-10-425-114-71125  
 ; Sequence 71125, Application US/10425114  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Liu, Jindong  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Screen, Steven E  
 ; APPLICANT: Tabaska, Jack E  
 ; APPLICANT: Cao, Yongwei  
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
 ; FILE REFERENCE: 38-21(5313)B  
 ; CURRENT APPLICATION NUMBER: US/10/425,114  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 73128  
 ; SEQ ID NO 71125  
 ; LENGTH: 159  
 ; TYPE: PRT  
 ; ORGANISM: Zea mays  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: LIB3597-011-F8\_FLI.pep  
 US-10-425-114-71125

Query Match 72.7%; Score 32; DB 12; Length 159;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 LRALGLK 9  
 |||||  
 Db 114 LRALGLK 120

Search completed: July 11, 2003, 12:35:09  
 Job time : 10.5574 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:54:32 ; Search time 62.9522 Seconds

(without alignments)  
204.833 Million cell updates/sec

Title: US-10-091-442-6

Perfect score: 98

Sequence: 1 ADLSGASGRLYSRMWK 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Pending\_Patents\_AA\_Main:\*

1: /cgn2\_6/ptodata/1/paa/US06.COMB.pep.\*  
2: /cgn2\_6/ptodata/1/paa/US06.COMB.pep.\*  
3: /cgn2\_6/ptodata/1/paa/US07.COMB.pep.\*  
4: /cgn2\_6/ptodata/1/paa/US08.COMB.pep.\*  
5: /cgn2\_6/ptodata/1/paa/US081.COMB.pep.\*  
6: /cgn2\_6/ptodata/1/paa/US081.COMB.pep.\*  
7: /cgn2\_6/ptodata/1/paa/US083.COMB.pep.\*  
8: /cgn2\_6/ptodata/1/paa/US084.COMB.pep.\*  
9: /cgn2\_6/ptodata/1/paa/US085.COMB.pep.\*  
10: /cgn2\_6/ptodata/1/paa/US086.COMB.pep.\*  
11: /cgn2\_6/ptodata/1/paa/US087.COMB.pep.\*  
12: /cgn2\_6/ptodata/1/paa/US088.COMB.pep.\*  
13: /cgn2\_6/ptodata/1/paa/US089.COMB.pep.\*  
14: /cgn2\_6/ptodata/1/paa/US090.COMB.pep.\*  
15: /cgn2\_6/ptodata/1/paa/US091.COMB.pep.\*  
16: /cgn2\_6/ptodata/1/paa/US092.COMB.pep.\*  
17: /cgn2\_6/ptodata/1/paa/US093.COMB.pep.\*  
18: /cgn2\_6/ptodata/1/paa/US094.COMB.pep.\*  
19: /cgn2\_6/ptodata/1/paa/US095.COMB.pep.\*  
20: /cgn2\_6/ptodata/1/paa/US096.COMB.pep.\*  
21: /cgn2\_6/ptodata/1/paa/US097.COMB.pep.\*  
22: /cgn2\_6/ptodata/1/paa/US098.COMB.pep.\*  
23: /cgn2\_6/ptodata/1/paa/US099.COMB.pep.\*  
24: /cgn2\_6/ptodata/1/paa/US100.COMB.pep.\*  
25: /cgn2\_6/ptodata/1/paa/US101.COMB.pep.\*  
26: /cgn2\_6/ptodata/1/paa/US102.COMB.pep.\*  
27: /cgn2\_6/ptodata/1/paa/US60.COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	98	100.0	20	15	US-09-140-719-6
2	98	100.0	20	24	US-10-091-442-6
3	92	93.9	20	4	US-08-091-028-6
4	92	93.9	20	4	US-08-091-028-6
5	90	91.8	78	27	US-60-160-203-4173
6	90	91.8	118	27	US-60-188-162-4539

7	90	91.8	380	4	US-08-091-028A-34	Sequence 34, Appl
8	90	91.8	380	15	US-09-140-719-34	Sequence 34, Appl
9	90	91.8	380	19	US-09-508-997A-2	Sequence 2, Appl
10	90	91.8	380	21	US-09-791-537-281	Sequence 281, App
11	90	91.8	380	23	US-09-936-883A-2	Sequence 2, Appl
12	90	91.8	380	24	US-10-091-442-34	Sequence 34, Appl
13	90	91.8	380	25	US-10-110-030-2	Sequence 2, Appl
14	83	84.7	368	19	US-09-508-997A-6	Sequence 6, Appl
15	83	84.7	368	23	US-09-936-883A-21	Sequence 21, Appl
16	83	84.7	368	25	US-10-110-030-6	Sequence 6, Appl
17	83	84.7	380	19	US-09-508-997A-4	Sequence 4, Appl
18	83	84.7	380	23	US-09-936-883A-19	Sequence 19, Appl
19	83	84.7	380	25	US-10-110-030-4	Sequence 4, Appl
20	54	55.1	195	21	US-09-791-537-102306	Sequence 102306,
21	53	54.1	405	21	US-09-791-537-99015	Sequence 99015, A
22	51	52.0	125	20	US-09-617-682A-3038	Sequence 3038, Ap
23	51	52.0	210	21	US-09-708-427-53024	Sequence 53024, A
24	51	52.0	210	26	US-10-219-999-47784	Sequence 47784, A
25	51	52.0	210	26	US-10-219-999-62585	Sequence 62585, A
26	51	52.0	210	26	US-10-312-544-9910	Sequence 9910, Ap
27	51	52.0	210	27	US-60-374-109-28769	Sequence 28769, A
28	51	52.0	306	21	US-09-708-427-53023	Sequence 53023, A
29	51	52.0	306	26	US-10-219-999-54475	Sequence 54475, A
30	51	52.0	306	26	US-10-219-999-54875	Sequence 54875, A
31	51	52.0	378	21	US-09-791-537-45859	Sequence 45859, A
32	51	52.0	378	24	US-10-023-634-84	Sequence 84, Appl
33	50	51.0	420	24	US-10-029-386-12104	Sequence 32104, A
34	50	51.0	615	1	PCT-US01-04942A-356	Sequence 356, App
35	50	51.0	615	20	US-09-629-469A-12768	Sequence 12768, A
36	50	51.0	615	21	US-09-791-537-112135	Sequence 112135,
37	50	51.0	615	26	US-10-221-278-356	Sequence 356, App
38	48	49.0	423	24	US-10-023-634-83	Sequence 83, Appl
39	48	49.0	591	21	US-09-791-537-132657	Sequence 132657,
40	47	48.0	94	27	US-60-186-718-6010	Sequence 6010, Ap
41	47	48.0	94	27	US-60-186-718-6011	Sequence 6011, Ap
42	47	48.0	365	24	US-10-023-634-20	Sequence 20, Appl
43	47	48.0	394	1	PCT-US01-13360A-59	Sequence 59, Appl
44	47	48.0	394	1	PCT-US01-13360A-59	Sequence 59, Appl
45	47	48.0	395	22	US-09-823-187-2	Sequence 2, Appl

#### ALIGNMENTS

RESULT 1  
US-09-140-719-6  
; Sequence 6, Application US/09140719  
; GENERAL INFORMATION:  
; APPLICANT: TSUJIMOTO, Masafumi  
; APPLICANT: IWASA, Fuyuki  
; APPLICANT: TSURUOKA, Nobuo  
; APPLICANT: NAKAZATO, Hiroshi  
; APPLICANT: MIURA, Kenju  
; APPLICANT: ISHIDA, Nobuhiko  
; APPLICANT: KURIHARA, Tatsuya  
; APPLICANT: YAMACHI, Koza  
; APPLICANT: YAMAGUCHI, Nozomi  
; TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
; NUMBER OF SEQUENCES: 34  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: P. O. Box 1404  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22133-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/140,719

FILED DATE: 08-AUG-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-140-719-6

Query Match 100.0%; Score 98; DB 15; Length 20;  
Best Local Similarity 100.0%; Pred. No. 1.4e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLSGIASGRLYISRMCK 20  
Db 1 ADLSGIASGRLYISRMCK 20

RESULT 2  
US-10-091-442-6  
Sequence 6, Application US/10091442  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
IMASA, Fuyuki  
TSURUOKA, Nobuo  
NAKAZATO, Hiroshi  
MURA, Kenju  
ISHIDA, Nobuhiro  
KURIHARA, Tatsuya  
YAMAGUCHI, Kozo  
YAMAGUCHI, Nozomi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/091,442  
FILING DATE: 07-Mar-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998

APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
US-10-091-442-6

Query Match 100.0%; Score 98; DB 24; Length 20;  
Best Local Similarity 100.0%; Pred. No. 1.4e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADLSGIASGRLYISRMCK 20  
Db 1 ADLSGIASGRLYISRMCK 20

RESULT 3  
US-08-091-028-6  
Sequence 6, Application US/08091028  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
IMASA, Fuyuki  
TSURUOKA, Nobuo  
NAKAZATO, Hiroshi  
MURA, Kenju  
ISHIDA, Nobuhiro  
KURIHARA, Tatsuya  
YAMAGUCHI, Kozo  
YAMAGUCHI, Nozomi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/091,028  
FILING DATE: 14-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Crane-Feury, Sharon E

```

:
:   REGISTRATION NUMBER: 36,113
:   REFERENCE/DOCKET NUMBER: 001560-204
:   TELECOMMUNICATION INFORMATION:
:   TELEPHONE: (703) 836-6620
:   TELEFAX: (703) 836-6620
:   INFORMATION FOR SEQ ID NO: 6:
:   SEQUENCE CHARACTERISTICS:
:   LENGTH: 20 amino acids
:   TYPE: amino acid
:   STRANDEDNESS: single
:   TOPOLOGY: linear
:   MOLECULE TYPE: peptide
:   US-08-091-028-6

Query Match      93.9%: Score 92; DB 4; Length 20;
Best Local Similarity 95.0%: Pred. No. 1.4e-07;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 ADLSGIASGRLYISRMWGK 20
        |||
Db      1 ADLSGIASGRLYISRMWGK 20

RESULT 4
US-08-091-028A-6
: Sequence 6, Application US/08091028A
: GENERAL INFORMATION:
:   APPLICANT: TSUJIMOTO, Masafumi
:   APPLICANT: IWASHI, Fuyuki
:   APPLICANT: TSURUOKA, Nobuo
:   APPLICANT: NAKAZATO, Hiroshi
:   APPLICANT: MURA, Kenji
:   APPLICANT: ISHIDA, Nobuhiko
:   APPLICANT: KURIHARA, Tatsuya
:   APPLICANT: YAMACHI, Kozo
:   APPLICANT: YAMAGUCHI, Nozomi
:   TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR
:   NUMBER OF SEQUENCES: 34
:   CORRESPONDENCE ADDRESSES:
:   ADDRESSEE: Burns, Doane, Swecker & Mathis
:   STREET: George Mason Bldg., Washington & Prince Sts.
:   CITY: Alexandria
:   STATE: Virginia
:   COUNTRY: United States
:   ZIP: 22313-1404
:   COMPUTER READABLE FORM:
:   MEDIUM TYPE: floppy disk
:   COMPUTER: IBM PC compatible
:   OPERATING SYSTEM: PC-DOS/MS-DOS
:   SOFTWARE: Patent Release #1.0, Version #1.25
:   CURRENT APPLICATION DATA:
:   APPLICATION NUMBER: US/08/091,028A
:   FILING DATE: 14-JUL-1993
:   CLASSIFICATION: 435
:   PRIOR APPLICATION DATA:
:   APPLICATION NUMBER: JP 4-212305
:   FILING DATE: 17-JUL-1992
:   PRIOR APPLICATION DATA:
:   APPLICATION NUMBER: JP 5-067339
:   FILING DATE: 04-MAR-1993
:   ATTORNEY/AGENT INFORMATION:
:   NAME: REA, TERESA STANER
:   REGISTRATION NUMBER: 30,427
:   REFERENCE/DOCKET NUMBER: 001560-204
:   TELECOMMUNICATION INFORMATION:
:   TELEPHONE: (703) 836-6620
:   TELEFAX: (703) 836-6620
:   INFORMATION FOR SEQ ID NO: 6:
:   SEQUENCE CHARACTERISTICS:
:   LENGTH: 20 amino acids
:   TYPE: amino acid
:   STRANDEDNESS: single
:   TOPOLOGY: linear

:
:   MOLECULE TYPE: peptide
:   US-08-091-028A-6

Query Match      93.9%: Score 92; DB 4; Length 20;
Best Local Similarity 95.0%: Pred. No. 1.4e-07;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 ADLSGIASGRLYISRMWGK 20
        |||
Db      1 ADLSGIASGRLYISRMWGK 20

RESULT 5
US-60-160-203-4173
: Sequence 4173, Application US/60160203
: GENERAL INFORMATION:
:   APPLICANT: BONAZZI, VIVIAN
:   TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
:   TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS AND
:   TITLE OF INVENTION: USES THEREOF
:   FILE REFERENCE: CLO00116
:   CURRENT APPLICATION NUMBER: US/60/160,203
:   CURRENT FILING DATE: 1999-10-19
:   NUMBER OF SEQ ID NOS: 6374
:   SOFTWARE: FASTSEQ for Windows Version 4.0
:   SEQ ID NO 4173
:   LENGTH: 78
:   TYPE: PRT
:   ORGANISM: HUMAN
:   US-60-160-203-4173

Query Match      91.8%: Score 90; DB 27; Length 78;
Best Local Similarity 95.0%: Pred. No. 1.5e-06;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 ADLSGIASGRLYISRMWGK 20
        |||
Db      58 ADLSGIASGRLYISRMWGK 77

RESULT 6
US-60-188-162-4539
: Sequence 4539, Application US/60188162
: GENERAL INFORMATION:
:   APPLICANT: Beasley, Ellen
:   TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
:   TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
:   TITLE OF INVENTION: USES THEREOF
:   FILE REFERENCE: CLO00518
:   CURRENT APPLICATION NUMBER: US/60/188,162
:   CURRENT FILING DATE: 2000-05-09
:   NUMBER OF SEQ ID NOS: 5094
:   SOFTWARE: FASTSEQ for Windows Version 4.0
:   SEQ ID NO 4539
:   LENGTH: 118
:   TYPE: PRT
:   ORGANISM: HUMAN
:   US-60-188-162-4539

Query Match      91.8%: Score 90; DB 27; Length 118;
Best Local Similarity 95.0%: Pred. No. 2.4e-06;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 ADLSGIASGRLYISRMWGK 20
        |||
Db      58 ADLSGIASGRLYISRMWGK 77

RESULT 7
US-08-091-028A-34
: Sequence 34, Application US/08091028A
: GENERAL INFORMATION:
:   APPLICANT: TSUJIMOTO, Masafumi
```

APPLICANT: IWASA, Fuyuki  
APPLICANT: TSURUOKA, Nobuo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, Nobuhiko  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, Nozomi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/091,028A  
FILING DATE: 14-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: REA, TERESA STANER  
REGISTRATION NUMBER: 30,427  
REFERENCE/DOCKET NUMBER: 001560-204  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-6620  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-091-028A-34

Query Match 91.8%; Score 90; DB 4; Length 380;  
Best Local Similarity 95.0%; Pred. No. 9.3e-06;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ADLSGIASGRLYISRMGK 20  
DB 305 ADLSGIASGRLYISRMHK 324

RESULT 8  
US-09-140-719-34  
Sequence 34, Application US/09140719  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSUROOKA, Nobuo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, Nobuhiko  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, Nozomi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Burns, Doane, Swecker & Mathis

STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-140-719-34

Query Match 91.8%; Score 90; DB 15; Length 380;  
Best Local Similarity 95.0%; Pred. No. 9.3e-06;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ADLSGIASGRLYISRMGK 20  
DB 305 ADLSGIASGRLYISRMHK 324

RESULT 9  
US-09-508-997A-2  
Sequence 2, Application US/09508997A  
GENERAL INFORMATION:  
APPLICANT: MIYATA, Toshio  
TITLE OF INVENTION: Megsin Protein  
FILE REFERENCE: SHIM-004  
CURRENT APPLICATION NUMBER: US/09/508,997A  
CURRENT FILING DATE: 2000-06-27  
PRIOR APPLICATION NUMBER: JP 09-275302  
PRIOR FILING DATE: 1997-09-22  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 2  
LENGTH: 380  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-508-997A-2

Query Match 91.8%; Score 90; DB 19; Length 380;  
Best Local Similarity 95.0%; Pred. No. 9.3e-06;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
OY      1 ADLSCIAGGRLYISRMGMK 20
DB      305 ADLSCIAGGRLYISRMGMK 324

RESULT 10
US-09-791-537-281
; Sequence 281, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 281
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-791-537-281

Query Match      91.8%; Score 90; DB 21; Length 380;
Best Local Similarity 95.0%; Pred. No. 9.3e-06;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      1 ADLSCIAGGRLYISRMGMK 20
DB      305 ADLSCIAGGRLYISRMGMK 324

RESULT 11
US-09-936-883A-2
; Sequence 2, Application US/09936883A
; GENERAL INFORMATION:
; APPLICANT: MIYATA, Toshio
; TITLE OF INVENTION: A Method for Detecting Megsin Protein and Use
; FILE REFERENCE: F2-101DPIPT
; CURRENT APPLICATION NUMBER: US/09/936,883A
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: JP 1999-75305
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: JP 1999-306623
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-936-883A-2

Query Match      91.8%; Score 90; DB 23; Length 380;
Best Local Similarity 95.0%; Pred. No. 9.3e-06;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      1 ADLSCIAGGRLYISRMGMK 20
DB      305 ADLSCIAGGRLYISRMGMK 324

RESULT 12
US-10-091-442-34
; Sequence 34, Application US/10091442
; GENERAL INFORMATION:
; APPLICANT: TSUJIMOTO, Masafumi
; APPLICANT: IWASA, Fuyuki
; TSUROUOKA, Nobuo
; NAKAZATO, Hiroshi

MIURA, Kenju
ISHIDA, Nobuhito
KURIHARA, Tatsuya
YAMACHI, KOZO
YAMAGUCHI, Nozomi
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESS: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/091,442
FILING DATE: 07-Mar-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/140,719
FILING DATE: 08-AUG-1998
APPLICATION NUMBER: US 08/474,661
FILING DATE: 07-JUN-1995
APPLICATION NUMBER: US 08/091,028
FILING DATE: 14-JUL-1993
APPLICATION NUMBER: JP 4-212305
FILING DATE: 17-JUL-1992
APPLICATION NUMBER: JP 6-067339
FILING DATE: 04-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: McGowan, Malcolm K.
REGISTRATION NUMBER: 39,300
REFERENCE/DOCKET NUMBER: 001560-247
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 380 amino acids
TYPE: amino acid
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 34:
US-10-091-442-34

Query Match      91.8%; Score 90; DB 24; Length 380;
Best Local Similarity 95.0%; Pred. No. 9.3e-06;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      1 ADLSCIAGGRLYISRMGMK 20
DB      305 ADLSCIAGGRLYISRMGMK 324

RESULT 13
US-10-110-030-2
; Sequence 2, Application US/10110030
; GENERAL INFORMATION:
; APPLICANT: Miyata, Toshio
; TITLE OF INVENTION: ANIMAL MODEL FOR MESENCHYMAL PROLIFERATIVE
; FILE REFERENCE: SHIM015
; CURRENT APPLICATION NUMBER: US/10/110,030
; CURRENT FILING DATE: 2002-04-03
; PRIOR APPLICATION NUMBER: 11/285736
; PRIOR FILING DATE: 1999-10-06
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
```

SEQ ID NO 2  
LENGTH: 380  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-110-030-2

Query Match 91.8%; Score 90; DB 25; Length 380;  
Best Local Similarity 95.0%; Pred. No. 9.3e-06;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ADLSGIASGRLYISRMGMK 20  
|||||  
Db 305 ADLSGIASGRLYISRMGMK 324

RESULT 14  
US-09-508-997A-6  
Sequence 6, Application US/09508997A  
GENERAL INFORMATION:  
APPLICANT: MIYATA, Toshio  
FILE REFERENCE: SHIM-004  
CURRENT APPLICATION NUMBER: US/09/508, 997A  
CURRENT FILING DATE: 2000-06-27  
PRIOR APPLICATION NUMBER: JP 09-275302  
PRIOR FILING DATE: 1997-09-22  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 6  
LENGTH: 368  
TYPE: PRT  
ORGANISM: Mus musculus  
US-09-508-997A-6

Query Match 84.7%; Score 83; DB 19; Length 368;  
Best Local Similarity 80.0%; Pred. No. 0.00013;  
Matches 16; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 ADLSGIASGRLYISRMGMK 20  
|||||  
Db 293 ADLSGIASGRLYISRMGMK 312

RESULT 15  
US-09-936-883A-21  
Sequence 21, Application US/09936883A  
GENERAL INFORMATION:  
APPLICANT: MIYATA, Toshio  
FILE REFERENCE: P2-101DPICT  
CURRENT APPLICATION NUMBER: US/09/936, 883A  
CURRENT FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: JP 1999-75305  
PRIOR FILING DATE: 1999-03-19  
PRIOR APPLICATION NUMBER: JP 1999-306623  
PRIOR FILING DATE: 1999-10-28  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 21  
LENGTH: 368  
TYPE: PRT  
ORGANISM: Mus musculus  
US-09-936-883A-21

Query Match 84.7%; Score 83; DB 23; Length 368;  
Best Local Similarity 80.0%; Pred. No. 0.00013;  
Matches 16; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 ADLSGIASGRLYISRMGMK 20  
|||||  
Db 293 ADLSGIASGRLYISRMGMK 312

Search completed: July 11, 2003, 12:27:34  
Job time : 63.9522 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: July 11, 2003, 11:55:27 ; Search time 18 7942 Seconds  
(without alignments)  
280.699 Million cell updates/sec

Title: US-10-091-442-6  
Perfect score: 98  
Sequence: 1 ADLSGIASGRLYISRMNGK 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1229476 seqs, 263775074 residues  
Total number of hits satisfying chosen parameters: 1229476

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

Pending\_Patents\_AA\_New:\*  
1: /cgn2\_6/ptodata/1/paa/PCT\_NEW\_COMB.pep4:\*  
2: /cgn2\_6/ptodata/1/paa/PCT\_NEW\_COMB.pep4:\*  
3: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pep4:\*  
4: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep4:\*  
5: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep4:\*  
6: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep4:\*  
7: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep4:\*  
8: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep4:\*  
9: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep4:\*  
10: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep4:\*  
11: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep4:\*  
12: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep4:\*  
13: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep4:\*  
14: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	90	91.8	380	10	US-09-936-883C-2
2	90	91.8	380	12	US-10-126-052A-263
3	90	91.8	380	14	US-60-443-566-3006
4	90	91.8	380	14	US-60-455-444-4692
5	90	91.8	380	14	US-60-465-241-4692
6	83	84.7	368	10	US-09-936-883C-21
7	83	84.7	380	10	US-09-936-883C-19
8	51	52.0	233	12	US-10-425-114-57272
9	51	52.0	233	12	US-10-425-114-68267
10	51	52.0	306	12	US-10-425-114-59731
11	51	52.0	306	12	US-10-425-114-59936
12	47	48.0	394	12	US-10-258-951-59
13	47	48.0	405	2	PCT-US02-38526-416
14	47	48.0	405	14	US-60-443-566-4407
15	47	48.0	405	14	US-60-455-444-7462
16	47	48.0	405	14	US-60-465-241-7462
17	47	48.0	425	2	PCT-US02-38526-408
18	47	48.0	425	2	PCT-US02-07215B-2
19	47	48.0	425	12	US-10-419-277-2

20	47	48.0	425	14	US-60-443-566-4406	Sequence 4406, Ap
21	47	48.0	425	14	US-60-455-444-7461	Sequence 7461, Ap
22	47	48.0	425	14	US-60-465-241-7461	Sequence 7461, Ap
23	47	48.0	617	2	PCT-US02-38526-415	Sequence 415, App
24	46	46.9	257	2	PCT-US02-40225-3054	Sequence 3054, Ap
25	46	46.9	257	12	US-10-320-797-3054	Sequence 3054, Ap
26	46	46.9	377	12	US-10-144-779-358	Sequence 358, App
27	45	45.9	238	10	US-09-724-676-67567	Sequence 67567, A
28	45	45.9	238	10	US-09-724-676-67568	Sequence 67568, A
29	45	45.9	238	10	US-09-724-676-67569	Sequence 67569, A
30	45	45.9	238	10	US-09-724-676-67570	Sequence 67570, A
31	45	45.9	238	10	US-09-724-676-67571	Sequence 67571, A
32	45	45.9	238	10	US-09-724-676-67572	Sequence 67572, A
33	45	45.9	238	10	US-09-724-676-67573	Sequence 67573, A
34	45	45.9	238	10	US-09-724-676-67574	Sequence 67574, A
35	45	45.9	238	10	US-09-724-676-67575	Sequence 67575, A
36	45	45.9	238	10	US-09-724-676-67576	Sequence 67576, A
37	45	45.9	238	10	US-09-724-676-67577	Sequence 67577, A
38	45	45.9	238	10	US-09-724-676-67578	Sequence 67578, A
39	45	45.9	238	10	US-09-724-676-67579	Sequence 67579, A
40	45	45.9	238	10	US-09-724-676-67580	Sequence 67580, A
41	45	45.9	238	10	US-09-724-676-67581	Sequence 67581, A
42	45	45.9	238	10	US-09-724-676-67582	Sequence 67582, A
43	45	45.9	238	10	US-09-724-676-67583	Sequence 67583, A
44	45	45.9	238	10	US-09-724-676-67584	Sequence 67584, A
45	45	45.9	238	10	US-09-724-676-67585	Sequence 67585, A

## ALIGNMENTS

RESULT 1  
US-09-936-883C-2  
; Sequence 2, Application US/09936883C  
; GENERAL INFORMATION:  
; APPLICANT: MIYATA, Toshio  
; TITLE OF INVENTION: A Method for Detecting Megsin Protein and Use  
; TITLE OF INVENTION: Thereof  
; FILE REFERENCE: P2-10DIPICST  
; CURRENT APPLICATION NUMBER: US/09/936,883C  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: JP 1999-75305  
; PRIOR FILING DATE: 1999-03-19  
; PRIOR APPLICATION NUMBER: JP 1999-306623  
; PRIOR FILING DATE: 1999-10-28  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 380  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-936-883C-2

Query Match 91.8%; Score 90; DB 10; Length 380;  
Best Local Similarity 95.0%; Pred. No. 1.3e-06;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ADLSGIASGRLYISRMNGK 20  
DB 305 ADLSGIASGRLYISRMNGK 324

RESULT 2  
US-10-126-052A-263  
; Sequence 263, Application US/10126052A  
; GENERAL INFORMATION:  
; APPLICANT: Aziz, Natasha  
; APPLICANT: Murray, Richard  
; APPLICANT: Eos Biotechnology, Inc.  
; TITLE OF INVENTION: Methods of Diagnosis of Lung Cancer, Compositions and  
; TITLE OF INVENTION: Methods of Screening for Modulators of Lung Cancer  
; FILE REFERENCE: 018501-001530US  
; CURRENT APPLICATION NUMBER: US/10/126,052A

```

: CURRENT FILING DATE: 2002-04-18
: PRIOR APPLICATION NUMBER: US 60/284,770
: PRIOR FILING DATE: 2001-04-18
: PRIOR APPLICATION NUMBER: US 60/290,492
: PRIOR FILING DATE: 2001-05-10
: PRIOR APPLICATION NUMBER: US 60/339,245
: PRIOR FILING DATE: 2001-11-09
: PRIOR APPLICATION NUMBER: US 60/350,666
: PRIOR FILING DATE: 2001-11-13
: PRIOR APPLICATION NUMBER: US 60/334,370
: PRIOR FILING DATE: 2001-11-29
: PRIOR APPLICATION NUMBER: US 60/372,246
: PRIOR FILING DATE: 2002-04-12
: NUMBER OF SEQ ID NOS: 691
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 263
: LENGTH: 380
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-126-052A-263
```

```

Query Match          91.8%; Score 90; DB 12; Length 380;
Best Local Similarity 95.0%; Pred. No. 1.3e-06;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 ADLSGIASGRLYISRMHGK 20
Db      305 ADLSGIASGRLYISRMHGK 324
```

```

RESULT 3
US-60-443-566-3006
: Sequence 3006, Application US/60443566
: GENERAL INFORMATION:
: APPLICANT: CARGILL, Michele
: APPLICANT: BEGOVICH, Ann
: TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
: FILE REFERENCE: C1001447
: CURRENT APPLICATION NUMBER: US/60/443,566
: CURRENT FILING DATE: 2003-01-30
: NUMBER OF SEQ ID NOS: 25102
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 3006
: LENGTH: 380
: TYPE: PRT
: ORGANISM: Homo sapiens
US-60-443-566-3006
```

```

Query Match          91.8%; Score 90; DB 14; Length 380;
Best Local Similarity 95.0%; Pred. No. 1.3e-06;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 ADLSGIASGRLYISRMHGK 20
Db      305 ADLSGIASGRLYISRMHGK 324
```

```

RESULT 4
US-60-455-444-4692
: Sequence 4692, Application US/60455444
: GENERAL INFORMATION:
: APPLICANT: CARGILL, Michele
: APPLICANT: BEGOVICH, Ann
: TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
: FILE REFERENCE: C1001455
: CURRENT APPLICATION NUMBER: US/60/455,444
: CURRENT FILING DATE: 2003-03-18
: NUMBER OF SEQ ID NOS: 50986
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 4692
: LENGTH: 380
```

```

: TYPE: PRT
: ORGANISM: Homo sapiens
US-60-455-444-4692
```

```

Query Match          91.8%; Score 90; DB 14; Length 380;
Best Local Similarity 95.0%; Pred. No. 1.3e-06;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 ADLSGIASGRLYISRMHGK 20
Db      305 ADLSGIASGRLYISRMHGK 324
```

```

RESULT 5
US-60-465-241-4692
: Sequence 4692, Application US/60465241
: GENERAL INFORMATION:
: APPLICANT: CARGILL, Michele
: APPLICANT: BEGOVICH, Ann
: TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
: FILE REFERENCE: C1001468
: CURRENT APPLICATION NUMBER: US/60/465,241
: CURRENT FILING DATE: 2003-04-23
: NUMBER OF SEQ ID NOS: 258418
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 4692
: LENGTH: 380
: TYPE: PRT
: ORGANISM: Homo sapiens
US-60-465-241-4692
```

```

Query Match          91.8%; Score 90; DB 14; Length 380;
Best Local Similarity 95.0%; Pred. No. 1.3e-06;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 ADLSGIASGRLYISRMHGK 20
Db      305 ADLSGIASGRLYISRMHGK 324
```

```

RESULT 6
US-09-936-883C-21
: Sequence 21, Application US/09936883C
: GENERAL INFORMATION:
: APPLICANT: MIYATA, Toshio
: TITLE OF INVENTION: A Method for Detecting Megsin Protein and Use
: FILE REFERENCE: P2-101DIPCT
: CURRENT APPLICATION NUMBER: US/09/936,883C
: CURRENT FILING DATE: 2001-12-21
: PRIOR APPLICATION NUMBER: JP 1999-75305
: PRIOR FILING DATE: 1999-03-19
: PRIOR APPLICATION NUMBER: JP 1999-306623
: PRIOR FILING DATE: 1999-10-28
: NUMBER OF SEQ ID NOS: 21
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 21
: LENGTH: 368
: TYPE: PRT
: ORGANISM: Mus musculus
US-09-936-883C-21
```

```

Query Match          84.7%; Score 83; DB 10; Length 368;
Best Local Similarity 80.0%; Pred. No. 1.9e-05;
Matches 16; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 ADLSGIASGRLYISRMHGK 20
Db      293 ADLSGIASGRLYISRMHGK 312
```

```
RESULT 7
```



```
US-09-936-883C-19
; Sequence 19, Application US/09936883C
; GENERAL INFORMATION:
; APPLICANT: MIYATA, Toshio
; TITLE OF INVENTION: A Method for Detecting Megsin Protein and Use
; FILE REFERENCE: P2-101DPCT
; CURRENT APPLICATION NUMBER: US/09/936, 883C
; PRIOR APPLICATION NUMBER: JP 1999-75305
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: JP 1999-306623
; PRIOR FILING DATE: 1999-10-28
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Rattus norvegicus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 51, 94
; OTHER INFORMATION: Xaa is unknown.
US-09-936-883C-19

Query Match      84.7%: Score 83; DB 10; Length 380;
Best Local Similarity 80.0%: Pred. No. 2e-05;
Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY      1 ADLSGIASGRLYISRM 20
Db      305 ADLSGIASGRLYISRM 324

RESULT 8
US-10-425-114-57272
; Sequence 57272, Application US/10425114
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovacic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425, 114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 57272
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB73109H11_FLI.pep
US-10-425-114-57272

Query Match      52.0%: Score 51; DB 12; Length 233;
Best Local Similarity 64.7%: Pred. No. 2.9;
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY      1 ADLSGIASGRLYISRM 17
Db      27 AGLSIESGRLYISNL 43

RESULT 9
US-10-425-114-68267
; Sequence 68267, Application US/10425114
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425, 114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 68267
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLMO17295D05_FLI.pep
US-10-425-114-68267

Query Match      52.0%: Score 51; DB 12; Length 233;
Best Local Similarity 64.7%: Pred. No. 2.9;
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY      1 ADLSGIASGRLYISRM 17
Db      27 AGLSIESGRLYISNL 43

RESULT 10
US-10-425-114-59731
; Sequence 59731, Application US/10425114
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovacic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425, 114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 59731
; LENGTH: 306
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3957-001-A2_FLI.pep
US-10-425-114-59731

Query Match      52.0%: Score 51; DB 12; Length 306;
Best Local Similarity 64.7%: Pred. No. 4;
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY      1 ADLSGIASGRLYISRM 17
Db      100 AGLSIESGRLYISNL 116

RESULT 11
US-10-425-114-59936
; Sequence 59936, Application US/10425114
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovacic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
```

;; CURRENT APPLICATION NUMBER: US/10/425,114  
;; CURRENT FILING DATE: 2003-04-28  
;; NUMBER OF SEQ ID NOS: 73128  
;; SEQ ID NO 59936  
;; LENGTH: 306  
;; TYPE: PRT  
;; ORGANISM: Zea mays  
;; FEATURE:  
;; OTHER INFORMATION: Clone ID: LIB3689-229-C3\_Flt.pep  
US-10-425-114-59936

Query Match 52.0%; Score 51; DB 12; Length 306;  
Best Local Similarity 64.7%; Pred. No. 4;  
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 ADLSGIASGRLYISRM 17  
DB 100 AGLSGIESGTYKLTISNL 116

RESULT 12  
US-10-258-951-59  
;; Sequence 59, Application US/10258951  
;; GENERAL INFORMATION:  
;; APPLICANT: Agarwal, Pankaj  
;; APPLICANT: Murodock, Paul R.  
;; APPLICANT: Rizvi, Safia K.  
;; APPLICANT: Smith, Randall F.  
;; APPLICANT: Xiang, Zhaoying  
;; APPLICANT: Kadinick, Karen  
;; APPLICANT: Lai, Ying-Ta  
;; APPLICANT: Xie, Qing  
;; TITLE OF INVENTION: NOVEL COMPOUNDS  
;; FILE REFERENCE: GP50025  
;; CURRENT APPLICATION NUMBER: US/10/258,951  
;; CURRENT FILING DATE: 2002-10-28  
;; PRIOR APPLICATION NUMBER: PCT/US01/13360  
;; PRIOR FILING DATE: 2001-04-26  
;; PRIOR APPLICATION NUMBER: 60/199,963  
;; PRIOR FILING DATE: 2000-04-27  
;; PRIOR APPLICATION NUMBER: 60/203,336  
;; PRIOR FILING DATE: 2000-05-11  
;; PRIOR APPLICATION NUMBER: 60/207,087  
;; PRIOR FILING DATE: 2000-05-25  
;; PRIOR APPLICATION NUMBER: 60/207,546  
;; PRIOR FILING DATE: 2000-05-26  
;; NUMBER OF SEQ ID NOS: 78  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 59  
;; LENGTH: 394  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-10-258-951-59

Query Match 48.0%; Score 47; DB 12; Length 394;  
Best Local Similarity 45.0%; Pred. No. 26;  
Matches 9; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

OY 1 ADLSGIASGRLYISRM 20  
DB 317 ADLTGISPSNLYLSKIIRK 336

RESULT 13  
PCT-US02-38526-416  
;; Sequence 416, Application PC/TUS0238526  
;; GENERAL INFORMATION:  
;; APPLICANT: Ghosh, Malabika  
;; APPLICANT: Tang, Y. Tom  
;; APPLICANT: Wang, Jian-Rui  
;; APPLICANT: Wang, Zhiwei  
;; APPLICANT: Zhao, Qing  
;; APPLICANT: Xu, Chongjun

;; APPLICANT: Mulero, Julio J  
;; APPLICANT: Boyle, Bryan J.  
;; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL POLYPEPTIDES AND POLYN  
;; FILE REFERENCE: HYS-B1/PCT  
;; CURRENT APPLICATION NUMBER: PCT/US02/38526  
;; CURRENT FILING DATE: 2002-12-03  
;; PRIOR APPLICATION NUMBER: US 09/488,725  
;; PRIOR FILING DATE: 2000-01-21  
;; PRIOR APPLICATION NUMBER: US 09/552,317  
;; PRIOR FILING DATE: 2000-04-25  
;; PRIOR APPLICATION NUMBER: PCT/US00/35017  
;; PRIOR FILING DATE: 2000-12-22  
;; PRIOR APPLICATION NUMBER: US 09/620,312  
;; PRIOR FILING DATE: 2000-07-19  
;; PRIOR APPLICATION NUMBER: PCT/US00/34263  
;; PRIOR FILING DATE: 2000-12-26  
;; PRIOR APPLICATION NUMBER: US 09/496,914  
;; PRIOR FILING DATE: 2000-02-03  
;; PRIOR APPLICATION NUMBER: US 09/560,875  
;; PRIOR FILING DATE: 2000-04-27  
;; PRIOR APPLICATION NUMBER: PCT/US01/03800  
;; PRIOR FILING DATE: 2001-02-05  
;; PRIOR APPLICATION NUMBER: US 09/598,075  
;; PRIOR FILING DATE: 2000-06-20  
;; PRIOR APPLICATION NUMBER: PCT/US01/04098  
;; PRIOR FILING DATE: 2001-02-05  
;; Remaining Prior Application data removed - See File Wrapper or PALM.  
;; NUMBER OF SEQ ID NOS: 653  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO 416  
;; LENGTH: 405  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
PCT-US02-38526-416

Query Match 48.0%; Score 47; DB 2; Length 405;  
Best Local Similarity 45.0%; Pred. No. 27;  
Matches 9; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

OY 1 ADLSGIASGRLYISRM 20  
DB 328 ADLTGISPSNLYLSKIIRK 347

RESULT 14  
US-60-443-566-4407  
;; Sequence 4407, Application US/60443566  
;; GENERAL INFORMATION:  
;; APPLICANT: CARGILL, Michele  
;; APPLICANT: BEGOVICH, Ann  
;; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
;; FILE REFERENCE: CL001447  
;; CURRENT APPLICATION NUMBER: US/60/443,566  
;; CURRENT FILING DATE: 2003-01-30  
;; NUMBER OF SEQ ID NOS: 25102  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 4407  
;; LENGTH: 405  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-60-443-566-4407

Query Match 48.0%; Score 47; DB 14; Length 405;  
Best Local Similarity 45.0%; Pred. No. 27;  
Matches 9; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

OY 1 ADLSGIASGRLYISRM 20  
DB 328 ADLTGISPSNLYLSKIIRK 347

RESULT 15

US-60-455-444-7462  
: Sequence 7462, Application US/60455444  
: GENERAL INFORMATION:  
: APPLICANT: CARGILL, Michele  
: APPLICANT: BEGOVICH, Ann  
: TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
: TITLE OF INVENTION: RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF  
: FILE REFERENCE: CLO01455  
: CURRENT APPLICATION NUMBER: US/60/455,444  
: CURRENT FILING DATE: 2003-03-18  
: NUMBER OF SEQ ID NOS: 50986  
: SOFTWARE: FastSeq for Windows Version 4.0  
: SEQ ID NO 7462  
: LENGTH: 405  
: TYPE: PRT  
: ORGANISM: Homo sapiens  
US-60-455-444-7462

Query Match 48.0%; Score 47; DB 14; Length 405;  
Best Local Similarity 45.0%; Pred. No. 27;  
Matches 9; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

OY 1 ADLSGIASGGRLYISRMGK 20  
|||:|: ||:::|  
Db 328 ADLTGISPSPLVLYLKIHK 347

Search completed: July 11, 2003, 12:35:10  
Job time : 19.8942 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:54:32 ; Search time 15.738 Seconds

(without alignments)  
204.833 Million cell updates/sec

Title: US-10-091-442-7

Perfect score: 26

Sequence: 1 LYDAK 5

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Pending\_Patents\_AA\_Main:\*

1: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/paa/US087\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/paa/US088\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/paa/US089\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/paa/US090\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/paa/US091\_COMB.pep.\*  
7: /cgn2\_6/ptodata/1/paa/US092\_COMB.pep.\*  
8: /cgn2\_6/ptodata/1/paa/US093\_COMB.pep.\*  
9: /cgn2\_6/ptodata/1/paa/US094\_COMB.pep.\*  
10: /cgn2\_6/ptodata/1/paa/US095\_COMB.pep.\*  
11: /cgn2\_6/ptodata/1/paa/US096\_COMB.pep.\*  
12: /cgn2\_6/ptodata/1/paa/US097\_COMB.pep.\*  
13: /cgn2\_6/ptodata/1/paa/US098\_COMB.pep.\*  
14: /cgn2\_6/ptodata/1/paa/US099\_COMB.pep.\*  
15: /cgn2\_6/ptodata/1/paa/US100\_COMB.pep.\*  
16: /cgn2\_6/ptodata/1/paa/US101\_COMB.pep.\*  
17: /cgn2\_6/ptodata/1/paa/US102\_COMB.pep.\*  
18: /cgn2\_6/ptodata/1/paa/US103\_COMB.pep.\*  
19: /cgn2\_6/ptodata/1/paa/US104\_COMB.pep.\*  
20: /cgn2\_6/ptodata/1/paa/US105\_COMB.pep.\*  
21: /cgn2\_6/ptodata/1/paa/US106\_COMB.pep.\*  
22: /cgn2\_6/ptodata/1/paa/US107\_COMB.pep.\*  
23: /cgn2\_6/ptodata/1/paa/US108\_COMB.pep.\*  
24: /cgn2\_6/ptodata/1/paa/US109\_COMB.pep.\*  
25: /cgn2\_6/ptodata/1/paa/US110\_COMB.pep.\*  
26: /cgn2\_6/ptodata/1/paa/US111\_COMB.pep.\*  
27: /cgn2\_6/ptodata/1/paa/US112\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	26	100.0	5	US-08-091-028-7
2	26	100.0	4	US-08-091-028A-7
3	26	100.0	5	US-09-140-719-7
4	26	100.0	15	US-10-091-442-7
5	26	100.0	5	US-60-196-718-3839
6	26	100.0	41	US-09-727-892-99

7	26	100.0	58	US-09-727-892A-99	Sequence 99, Appl
8	26	100.0	75	US-60-160-209-2854	Sequence 2854, Ap
9	26	100.0	75	US-60-163-062-871	Sequence 871, App
10	26	100.0	76	US-09-758-471-3660	Sequence 3660, Ap
11	26	100.0	76	US-10-235-953-3660	Sequence 3660, Ap
12	26	100.0	81	US-09-540-209B-7383	Sequence 7383, Ap
13	26	100.0	99	US-60-160-203-3968	Sequence 3968, Ap
14	26	100.0	99	US-60-163-123-1299	Sequence 1299, Ap
15	26	100.0	114	US-60-162-247-3259	Sequence 3259, Ap
16	26	100.0	117	US-09-791-537-71302	Sequence 71302, A
17	26	100.0	118	PCT-US00-07661-169	Sequence 169, App
18	26	100.0	118	PCT-US00-07661-171	Sequence 171, App
19	26	100.0	118	PCT-US00-07661-169	Sequence 169, App
20	26	100.0	118	PCT-US00-07661-171	Sequence 171, App
21	26	100.0	133	US-10-015-127-9753	Sequence 9753, Ap
22	26	100.0	151	US-09-513-996A-33705	Sequence 33705, A
23	26	100.0	151	US-09-513-996A-33704	Sequence 33704, A
24	26	100.0	155	US-09-513-996A-33704	Sequence 33704, A
25	26	100.0	155	US-09-688-051-891	Sequence 891, App
26	26	100.0	167	US-09-614-450-5014	Sequence 5014, Ap
27	26	100.0	171	US-09-614-450-5013	Sequence 5013, Ap
28	26	100.0	196	US-08-827-356-2912	Sequence 2912, Ap
29	26	100.0	199	US-09-791-537-147392	Sequence 147392, A
30	26	100.0	209	US-60-360-039-17822	Sequence 17822, A
31	26	100.0	211	US-09-675-784A-7278	Sequence 7278, Ap
32	26	100.0	212	PCT-US01-08631-55189	Sequence 55189, A
33	26	100.0	214	US-60-140-956-1585	Sequence 1585, Ap
34	26	100.0	215	US-60-167-217-8763	Sequence 8763, Ap
35	26	100.0	216	US-09-540-209B-5974	Sequence 5974, Ap
36	26	100.0	221	US-09-770-767-57481	Sequence 57481, A
37	26	100.0	221	US-09-270-849B-182166	Sequence 182166, A
38	26	100.0	227	US-09-611-529-4520	Sequence 4520, Ap
39	26	100.0	263	US-09-513-996A-29703	Sequence 29703, A
40	26	100.0	263	US-09-513-996A-46566	Sequence 46566, A
41	26	100.0	263	US-09-689-980-341	Sequence 341, App
42	26	100.0	281	US-09-791-537-115423	Sequence 115423, A
43	26	100.0	282	US-09-739-449-8924	Sequence 8924, Ap
44	26	100.0	282	US-09-803-110-8924	Sequence 8924, Ap
45	26	100.0	291	US-10-219-999-47475	Sequence 47475, A

## ALIGNMENTS

RESULT 1  
US-08-091-028-7  
Sequence 7, Application US/08091028  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IMASA, Fuyuki  
APPLICANT: TSURUOKA, Nobuo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, Nobuhiro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAGUCHI, Kozo  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/091,028

FILING DATE: 14-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Crane-Feury, Sharon E  
REGISTRATION NUMBER: 36,113  
REFERENCE/DOCKET NUMBER: 001560-204  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-6620  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-091-028-7

Query Match 100.0%; Score 26; DB 4; Length 5;  
Best Local Similarity 100.0%; Pred. No. 4.2e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5  
DB 1 LYDAK 5

## RESULT 2

US-08-091-028A-7  
Sequence 7, Application US/08091028A  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSURUOKA, Nobuo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, Nobuhiko  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAIICHI, Kozo  
APPLICANT: YAMAGUCHI, Nozomi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/091,028A  
FILING DATE: 14-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: REA, TERESA STANER  
REGISTRATION NUMBER: 30,427

REFERENCE/DOCKET NUMBER: 001560-204  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-6620  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-091-028A-7

Query Match 100.0%; Score 26; DB 4; Length 5;  
Best Local Similarity 100.0%; Pred. No. 4.2e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5  
DB 1 LYDAK 5

## RESULT 3

US-09-140-719-7  
Sequence 7, Application US/09140719  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSURUOKA, Nobuo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, Nobuhiko  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAIICHI, Kozo  
APPLICANT: YAMAGUCHI, Nozomi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:  
LENGTH: 5 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-140-719-7

Query Match 100.0%; Score 26; DB 15; Length 5;  
Best Local Similarity 100.0%; Pred. No. 4.2e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5  
|||||  
Db 1 LYDAK 5

#### RESULT 4

US-10-091-442-7  
Sequence 7, Application US/10091442

##### GENERAL INFORMATION:

APPLICANT: TSUJIMOTO, Masafumi

IMASA, Fuyuki

TSUROOKA, Nobuo

NAKAZATO, Hiroshi

MURA, Kenju

ISHIDA, Nobuhito

KURIHARA, Tatsuya

YAMAGUCHI, Kozo

YAMAGUCHI, Nozomi

TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR

NUMBER OF SEQUENCES: 34

CORRESPONDENCE ADDRESS:

ADDRESSEE: Burns, Doane, Swecker & Mathis

STREET: P.O. Box 1404

CITY: Alexandria

STATE: Virginia

COUNTRY: United States

ZIP: 22313-1404

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/091,442

FILING DATE: 07-Mar-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/140,719

FILING DATE: 08-Aug-1998

APPLICATION NUMBER: US 08/474,661

FILING DATE: 07-JUN-1995

APPLICATION NUMBER: US 08/091,028

FILING DATE: 14-JUL-1993

APPLICATION NUMBER: JP 4-212305

FILING DATE: 17-JUL-1992

APPLICATION NUMBER: JP 6-067339

ATTORNEY/AGENT INFORMATION:

NAME: McGowan, Malcolm K.

REGISTRATION NUMBER: 39,300

TELECOMMUNICATION INFORMATION:

REFERENCE/DOCKET NUMBER: 001560-247

TELEPHONE: (703) 836-6620

TELEFAX: (703) 836-2021

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 5 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 7:  
US-10-091-442-7

Query Match 100.0%; Score 26; DB 24; Length 5;  
Best Local Similarity 100.0%; Pred. No. 4.2e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5  
|||||  
Db 1 LYDAK 5

#### RESULT 5

US-60-196-718-3839  
Sequence 3839, Application US/60196718

##### GENERAL INFORMATION:

APPLICANT: Bonazzi, Vivien

TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,

TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND

TITLE OF INVENTION: USFS THEREOF

FILE REFERENCE: C1000456

CURRENT APPLICATION NUMBER: US/60/196,718

CURRENT FILING DATE: 2000-04-13

NUMBER OF SEQ ID NOS: 7494

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 3839

LENGTH: 41

TYPE: PRT

ORGANISM: HUMAN

FEATURE:

NAME/KEY: VARIANT

LOCATION: (1)...(41)

OTHER INFORMATION: Xaa = Any Amino Acid

Query Match 100.0%; Score 26; DB 27; Length 41;  
Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5  
|||||  
Db 10 LYDAK 14

#### RESULT 6

US-09-727-892-99  
Sequence 99, Application US/09727892

##### GENERAL INFORMATION:

APPLICANT: PhageTech, Inc.

APPLICANT: PELLETIER, Jerry

APPLICANT: GROS, Philippe

APPLICANT: DOBOW, Michael

TITLE OF INVENTION: DNA SEQUENCES FROM STAPHYLOCOCCUS AUREUS BACTERIOPHAGE 44 AHJD

TITLE OF INVENTION: THAT ENCODE ANTI-MICROBIAL POLYPEPTIDES

FILE REFERENCE: 073406-0302

CURRENT APPLICATION NUMBER: US/09/727,892

CURRENT FILING DATE: 2000-12-01

NUMBER OF SEQ ID NOS: 159

SOFTWARE: Patentin version 3.0

SEQ ID NO 99

LENGTH: 58

TYPE: PRT

ORGANISM: Staphylococcus aureus Bacteriophage 44 AHJD

US-09-727-892-99

Query Match 100.0%; Score 26; DB 21; Length 58;  
Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5  
|||||  
Db 29 LYDAK 33

```
RESULT 7
US-09-727-892A-99
; Sequence 99, Application US/09727892A
; GENERAL INFORMATION:
; APPLICANT: PELLETIER, JERRY
; APPLICANT: GROS, PHILIPPE
; TITLE OF INVENTION: DNA SEQUENCES FROM STAPHYLOCOCCUS AUREUS BACTERIOPHAGE 44AHJD
; FILE REFERENCE: 073406/0302
; CURRENT APPLICATION NUMBER: US/09/727, 892A
; CURRENT FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: 60/168, 777
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 170
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 99
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Staphylococcus aureus Bacteriophage 44AHJD
US-09-727-892A-99

Query Match
Best Local Similarity 100.0%; Score 26; DB 21; Length 58;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5
DB 29 LYDAK 33

RESULT 8
US-60-160-209-2854
; Sequence 2854, Application US/60160209
; GENERAL INFORMATION:
; APPLICANT: BONAZZI, VIVIAN
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS AND USES
; FILE REFERENCE: C1000113
; CURRENT APPLICATION NUMBER: US/60/160, 209
; CURRENT FILING DATE: 1999-10-19
; NUMBER OF SEQ ID NOS: 4646
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2854
; LENGTH: 75
; TYPE: PRT
; ORGANISM: HUMAN
US-60-160-209-2854

Query Match
Best Local Similarity 100.0%; Score 26; DB 27; Length 75;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5
DB 33 LYDAK 37

RESULT 9
US-60-163-062-871
; Sequence 871, Application US/60163062
; GENERAL INFORMATION:
; APPLICANT: BONAZZI, VIVIAN
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS AND USES
; FILE REFERENCE: C1000134
; CURRENT APPLICATION NUMBER: US/60/163, 062
; CURRENT FILING DATE: 1999-11-02
; NUMBER OF SEQ ID NOS: 1302
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 871
; LENGTH: 75
; TYPE: PRT
; ORGANISM: Human
US-60-163-062-871

Query Match
Best Local Similarity 100.0%; Score 26; DB 27; Length 75;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5
DB 33 LYDAK 37

RESULT 10
US-09-758-471-3660
; Sequence 3660, Application US/09758471
; GENERAL INFORMATION:
; APPLICANT: ROSEN ET AL.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PH002
; CURRENT APPLICATION NUMBER: US/09/758, 471
; CURRENT FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/179, 065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180, 628
; PRIOR FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 5386
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3660
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-758-471-3660

Query Match
Best Local Similarity 100.0%; Score 26; DB 21; Length 76;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5
DB 72 LYDAK 76

RESULT 11
US-10-235-953-3660
; Sequence 3660, Application US/10235953
; GENERAL INFORMATION:
; APPLICANT: ROSEN ET AL.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PH002CIN
; CURRENT APPLICATION NUMBER: US/10/235, 953
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: 09/758, 471
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/179, 065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180, 628
; PRIOR FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 5386
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3660
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-235-953-3660

Query Match
Best Local Similarity 100.0%; Score 26; DB 26; Length 76;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5
```



Db 72 LYDAK 76

RESULT 12

US-09-540-2098-7383

Sequence 7383, Application US/09540209B

GENERAL INFORMATION:

APPLICANT: Gary L. Breton

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTERIOIDES FR

TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 2709.1001-001

CURRENT APPLICATION NUMBER: US/09/540, 209B

CURRENT FILING DATE: 2000-04-04

NUMBER OF SEQ ID NOS: 10444

SEQ ID NO 7383

LENGTH: 81

TYPE: PRT

ORGANISM: B.fragilis

US-09-540-2098-7383

Query Match  
Best Local Similarity 100.0%; Score 26; DB 19; Length 81;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5

Db 69 LYDAK 73

RESULT 13

US-60-160-203-3968

Sequence 3968, Application US/60160203

GENERAL INFORMATION:

APPLICANT: BONAZZI, VIVIAN

TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,

TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS AND

FILE REFERENCE: CLO000116

CURRENT APPLICATION NUMBER: US/60/160, 203

CURRENT FILING DATE: 1999-10-19

NUMBER OF SEQ ID NOS: 6374

SOFTWARE: FASTSEQ for Windows Version 4.0

SEQ ID NO 3968

LENGTH: 99

TYPE: PRT

ORGANISM: HUMAN

US-60-160-203-3968

Query Match  
Best Local Similarity 100.0%; Score 26; DB 27; Length 99;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5

Db 50 LYDAK 54

RESULT 14

US-60-163-123-1299

Sequence 1299, Application US/60163123

GENERAL INFORMATION:

APPLICANT: BONAZZI, VIVIAN

TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,

TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS AND

FILE REFERENCE: CLO00137

CURRENT APPLICATION NUMBER: US/60/163, 123

CURRENT FILING DATE: 1999-11-02

NUMBER OF SEQ ID NOS: 1986

SOFTWARE: FASTSEQ for Windows Version 4.0

SEQ ID NO 1299

LENGTH: 99

TYPE: PRT  
ORGANISM: Human  
US-60-163-123-1299

Query Match  
Best Local Similarity 100.0%; Score 26; DB 27; Length 99;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5

Db 50 LYDAK 54

RESULT 15

US-60-162-247-3259

Sequence 3259, Application US/60162247

GENERAL INFORMATION:

APPLICANT: BONAZZI, VIVIAN

TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,

TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS AND

FILE REFERENCE: CLO00127

CURRENT APPLICATION NUMBER: US/60/162, 247

CURRENT FILING DATE: 1999-10-29

NUMBER OF SEQ ID NOS: 5442

SOFTWARE: FASTSEQ for Windows Version 4.0

SEQ ID NO 3259

LENGTH: 114

TYPE: PRT

ORGANISM: Human

US-60-162-247-3259

Query Match  
Best Local Similarity 100.0%; Score 26; DB 27; Length 114;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYDAK 5

Db 65 LYDAK 69

Search completed: July 11, 2003, 12:27:36  
Job time: 17.738 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:55:27 ; Search time 4.69854 Seconds

(without alignments)  
280.699 Million cell updates/sec

Title: US-10-091-442-7

Perfect score: 26

Sequence: 1 LYDAK 5

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1229476 seqs, 26375074 residues

Total number of hits satisfying chosen parameters: 1229476

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending Patents\_AA.New:\*  
1: /cgn2\_6/ptodata/1/paa/PCT\_NEW\_COMB.pep4:\*  
2: /cgn2\_6/ptodata/1/paa/PCT\_NEW\_COMB.pep4:\*  
3: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pep4:\*  
4: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pep4:\*  
5: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep4:\*  
6: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep4:\*  
7: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep4:\*  
8: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep4:\*  
9: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep4:\*  
10: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep4:\*  
11: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep4:\*  
12: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep4:\*  
13: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep4:\*  
14: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	26	100.0	209	12 US-10-369-493-17822	Sequence 17822, A
2	26	100.0	211	12 US-09-675-784A-7278	Sequence 7278, Ap
3	26	100.0	213	12 US-10-437-963-115384	Sequence 115384, A
4	26	100.0	227	12 US-09-950-084-4520	Sequence 4520, Ap
5	26	100.0	253	12 US-10-282-122A-60526	Sequence 60526, A
6	26	100.0	257	12 US-10-424-599-219617	Sequence 219617, A
7	26	100.0	265	12 US-10-437-963-144912	Sequence 144912, A
8	26	100.0	291	12 US-10-425-114-55804	Sequence 55804, A
9	26	100.0	319	12 US-10-434-437-98	Sequence 98, Appl
10	26	100.0	333	12 US-09-724-676-56008	Sequence 56008, A
11	26	100.0	343	12 US-09-724-676A-56008	Sequence 56008, A
12	26	100.0	345	12 US-10-366-683-21042	Sequence 21042, A
13	26	100.0	380	12 US-10-419-128-21042	Sequence 21042, A
14	26	100.0	380	12 US-09-936-883C-2	Sequence 2, Appl
15	26	100.0	380	12 US-10-126-052A-263	Sequence 263, App
16	26	100.0	380	14 US-60-443-566-3006	Sequence 3006, Ap
17	26	100.0	380	14 US-60-445-444-4692	Sequence 4692, Ap
18	26	100.0	380	14 US-60-465-241-4692	Sequence 4692, Ap
19	26	100.0	395	12 US-10-282-122A-48743	Sequence 48743, A

20	26	100.0	415	10 US-09-134-000C-3595	Sequence 3595, Ap
21	26	100.0	415	10 US-09-134-000C-3595	Sequence 3595, Ap
22	26	100.0	415	12 US-10-434-665-3595	Sequence 3595, Ap
23	26	100.0	427	12 US-10-454-437-96	Sequence 96, Appl
24	26	100.0	430	12 US-10-156-761-14800	Sequence 14800, A
25	26	100.0	454	12 US-10-310-154-660	Sequence 660, App
26	26	100.0	458	12 US-10-446-203-11632	Sequence 11632, A
27	26	100.0	489	2 PCT-US02-38445-42	Sequence 42, Appl
28	26	100.0	491	12 US-10-417-886-9542	Sequence 9542, Ap
29	26	100.0	496	12 US-10-072-012-771	Sequence 771, App
30	26	100.0	516	12 US-10-437-963-190810	Sequence 190810, A
31	26	100.0	521	12 US-10-072-012-769	Sequence 769, App
32	26	100.0	524	12 US-10-369-493-20966	Sequence 20966, A
33	26	100.0	529	12 US-10-431-652-6406	Sequence 6406, Ap
34	26	100.0	558	12 US-10-369-493-10997	Sequence 10997, A
35	26	100.0	559	12 US-10-431-652-552	Sequence 552, Ap
36	26	100.0	567	12 US-10-156-761-7781	Sequence 7781, Ap
37	26	100.0	574	12 US-10-369-493-23086	Sequence 23086, A
38	26	100.0	587	12 US-10-369-493-19378	Sequence 19378, A
39	26	100.0	588	12 US-10-366-683-24185	Sequence 24185, A
40	26	100.0	588	12 US-10-419-128-24185	Sequence 24185, A
41	26	100.0	597	12 US-10-282-122A-62084	Sequence 62084, A
42	26	100.0	625	12 US-10-425-114-58457	Sequence 58457, A
43	26	100.0	692	12 US-10-282-122A-72270	Sequence 72270, A
44	26	100.0	707	12 US-10-156-761-7899	Sequence 7899, Ap
45	26	100.0	712	12 US-10-156-761-15026	Sequence 15026, A

#### ALIGNMENTS

RESULT 1  
US-10-369-493-17822  
; Sequence 17822, Application US/10369493  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 17822  
; LENGTH: 209  
; TYPE: PRT  
; ORGANISM: SPHINGOMONAS  
US-10-369-493-17822

Query Match 100.0%; Score 26; DB 12; Length 209;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LYDAK 5  
Db 128 LYDAK 132

RESULT 2  
US-09-675-784A-7278  
; Sequence 7278, Application US/09675784A  
; GENERAL INFORMATION:  
; APPLICANT: HARE, ROBERTA S.  
; APPLICANT: SHAW, KAREN J.  
; APPLICANT: SHIMER JR., GEORGE H.  
; APPLICANT: KESSLER, MARCO  
; APPLICANT: NOLLING, JORK  
; APPLICANT: ZENG, QIANDONG

```

; APPLICANT: GREENE, JONATHAN R.
; TITLE OF INVENTION: ASPERGILLUS FUMIGATUS NUCLEIC ACIDS AND POLYPEPTIDES,
; FILE REFERENCE: 2976-4020051
; CURRENT APPLICATION NUMBER: US/09/675,784A
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/156,338
; NUMBER OF SEQ ID NOS: 13925
; SEQ ID NO 7278
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-09-675-784A-7278

Query Match
Best Local Similarity 100.0%; Score 26; DB 10; Length 211;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5
Db 206 LYDAK 210

RESULT 3
US-10-437-963-115384
; Sequence 115384, Application US/10437963
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 115384
; LENGTH: 213
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_18987C.1.pcp
US-10-437-963-115384

Query Match
Best Local Similarity 100.0%; Score 26; DB 12; Length 213;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5
Db 67 LYDAK 71

RESULT 4
US-09-950-084-4520
; Sequence 4520, Application US/09950084
; GENERAL INFORMATION:
; APPLICANT: George H. Shimer, Jr.
; APPLICANT: George H. Miller
; APPLICANT: Roberta S. Hare
; APPLICANT: Karen J. Shaw
; TITLE OF INVENTION: Staphylococcus aureus Related Compositions and Methods
; FILE REFERENCE: 1034/1C9630S2
; CURRENT APPLICATION NUMBER: US/09/950,084
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: US 09/417,811
; PRIOR FILING DATE: 1999-10-14
```

```

; PRIOR APPLICATION NUMBER: US 09/353,718
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: US 09/266,557
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 09/266,556
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 09/266,555
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 09/266,542
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 09/266,541
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 09/037,934
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: US 09/036,720
; PRIOR FILING DATE: 1998-03-06
; PRIOR APPLICATION NUMBER: US 09/036,338
; PRIOR FILING DATE: 1998-03-06
; Remaining Prior Application data removed - See file Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 7451
; SEQ ID NO 4520
; LENGTH: 227
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-950-084-4520

Query Match
Best Local Similarity 100.0%; Score 26; DB 10; Length 227;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5
Db 128 LYDAK 132

RESULT 5
US-10-282-122A-60526
; Sequence 60526, Application US/10282122A
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA_03A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
```

```

; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 60526
; LENGTH: 253
; TYPE: PRF
; ORGANISM: Listeria monocytogenes
; US-10-282-122a-60526

Query Match
Best Local Similarity 100.0%; Score 26; DB 12; Length 253;
Pred. No. 3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5
Db 231 LYDAK 235

RESULT 6
; US-10-424-599-219617
; Sequence 219617, Application US/10424599
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 219617
; LENGTH: 257
; TYPE: PRF
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_40342C.1.pep
; US-10-424-599-219617

Query Match
Best Local Similarity 100.0%; Score 26; DB 12; Length 257;
Pred. No. 3.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5
Db 239 LYDAK 243

RESULT 7
; US-10-437-963-144912
; Sequence 144912, Application US/10437963
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbasuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 144912
; LENGTH: 265
; TYPE: PRF
; ORGANISM: Oryza sativa
```

```

; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)-(265)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_45682C.1.pep
; US-10-437-963-144912

Query Match
Best Local Similarity 100.0%; Score 26; DB 12; Length 265;
Pred. No. 3.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5
Db 45 LYDAK 49

RESULT 8
; US-10-425-114-55804
; Sequence 55804, Application US/10425114
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 55804
; LENGTH: 291
; TYPE: PRF
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-GMROPIC033E03_FLI.pep
; US-10-425-114-55804

Query Match
Best Local Similarity 100.0%; Score 26; DB 12; Length 291;
Pred. No. 3.4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5
Db 34 LYDAK 38

RESULT 9
; US-10-454-437-98
; Sequence 98, Application US/10454437
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberhauser, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS
; FILE REFERENCE: BGI-1286PCN
; CURRENT APPLICATION NUMBER: US/10/454,437
; CURRENT FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: US 60/141031
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: DE 19931636.8
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19932125.6
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932126.4
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932127.2
```

;; PRIOR FILING DATE: 1999-07-09  
;; PRIOR APPLICATION NUMBER: DE 19932128.0  
;; PRIOR FILING DATE: 1999-07-09  
;; PRIOR APPLICATION NUMBER: DE 19932129.9  
;; PRIOR FILING DATE: 1999-07-19  
;; PRIOR APPLICATION NUMBER: DE 19932226.0  
;; PRIOR FILING DATE: 1999-07-09  
;; PRIOR APPLICATION NUMBER: DE 19932920.6  
;; PRIOR FILING DATE: 1999-07-14  
;; PRIOR APPLICATION NUMBER: DE 19932922.2  
;; PRIOR FILING DATE: 1999-07-14  
;; Remaining Prior Application data removed - See File Wrapper or PALM.  
;; NUMBER OF SEQ ID NOS: 442  
;; SEQ ID NO 98  
;; LENGTH: 319  
;; TYPE: PRT  
;; ORGANISM: Corynebacterium glutamicum  
US-10-454-437-98

Query Match 100.0%; Score 26; DB 12; Length 319;  
Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5  
|||||  
DB 246 LYDAK 250

RESULT 10  
US-09-724-676-56008  
;; Sequence 56008, Application US/09724676  
;; GENERAL INFORMATION:  
;; APPLICANT: Compugen LTD  
;; TITLE OF INVENTION: Variants of alternative splicing  
;; FILE REFERENCE: 129181.4 Compugen  
;; CURRENT APPLICATION NUMBER: US/09/724,676  
;; CURRENT FILING DATE: 2000-11-28  
;; NUMBER OF SEQ ID NOS: 97222  
;; SOFTWARE: PatentIn version 3.2  
;; SEQ ID NO 56008  
;; LENGTH: 333  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-09-724-676-56008

Query Match 100.0%; Score 26; DB 10; Length 333;  
Best Local Similarity 100.0%; Pred. No. 3.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5  
|||||  
DB 33 LYDAK 37

RESULT 11  
US-09-724-676A-56008  
;; Sequence 56008, Application US/09724676A  
;; GENERAL INFORMATION:  
;; APPLICANT: Compugen LTD  
;; TITLE OF INVENTION: Variants of alternative splicing  
;; FILE REFERENCE: 129181.4 Compugen  
;; CURRENT APPLICATION NUMBER: US/09/724,676A  
;; CURRENT FILING DATE: 2000-11-28  
;; NUMBER OF SEQ ID NOS: 97222  
;; SOFTWARE: PatentIn version 3.2  
;; SEQ ID NO 56008  
;; LENGTH: 333  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-09-724-676A-56008

Query Match 100.0%; Score 26; DB 10; Length 333;  
Best Local Similarity 100.0%; Pred. No. 3.9e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 LYDAK 5  
|||||  
DB 33 LYDAK 37

RESULT 12  
US-10-366-683-21042  
;; Sequence 21042, Application US/10366683  
;; GENERAL INFORMATION:  
;; APPLICANT: Rubenfield, Marc J.  
;; APPLICANT: Nolling, Jork  
;; APPLICANT: Deloughery, Craig  
;; APPLICANT: Bush, David  
;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
;; FILE REFERENCE: PATH03-04  
;; CURRENT APPLICATION NUMBER: US/10/366,683  
;; CURRENT FILING DATE: 2003-02-13  
;; PRIOR APPLICATION NUMBER: 09/252,991  
;; PRIOR FILING DATE: 1999-02-18  
;; NUMBER OF SEQ ID NOS: 33142  
;; SEQ ID NO 21042  
;; LENGTH: 345  
;; TYPE: PRT  
;; ORGANISM: Pseudomonas aeruginosa  
US-10-366-683-21042

Query Match 100.0%; Score 26; DB 12; Length 345;  
Best Local Similarity 100.0%; Pred. No. 4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5  
|||||  
DB 331 LYDAK 335

RESULT 13  
US-10-419-128-21042  
;; Sequence 21042, Application US/10419128  
;; GENERAL INFORMATION:  
;; APPLICANT: Marc J. Rubenfield et al.  
;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
;; FILE REFERENCE: 107196.136  
;; CURRENT APPLICATION NUMBER: US/10/419,128  
;; CURRENT FILING DATE: 2003-04-21  
;; PRIOR APPLICATION NUMBER: US/09/252,991  
;; PRIOR FILING DATE: 1999-02-18  
;; PRIOR APPLICATION NUMBER: US 60/074,788  
;; PRIOR FILING DATE: 1998-02-18  
;; PRIOR APPLICATION NUMBER: US 60/094,190  
;; PRIOR FILING DATE: 1998-07-27  
;; NUMBER OF SEQ ID NOS: 33142  
;; SEQ ID NO 21042  
;; LENGTH: 345  
;; TYPE: PRT  
;; ORGANISM: Pseudomonas aeruginosa  
US-10-419-128-21042

Query Match 100.0%; Score 26; DB 12; Length 345;  
Best Local Similarity 100.0%; Pred. No. 4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYDAK 5  
|||||  
DB 331 LYDAK 335

RESULT 14  
US-09-936-883C-2  
;; Sequence 2, Application US/09936883C

```
; GENERAL INFORMATION:
; APPLICANT: MIYATA, Toshio
; TITLE OF INVENTION: A Method for Detecting Megsin Protein and Use
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: F2-101DPICT
; CURRENT APPLICATION NUMBER: US/09/936, 883C
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: JP 1999-75305
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: JP 1999-306623
; PRIOR FILING DATE: 1999-10-28
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-936-883C-2
```

```
Query Match          100.0%; Score 26; DB 10; Length 380;
Best Local Similarity 100.0%; Pred No. 4.4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 LYDAK 5
        |||||
Db       121 LYDAK 125
```

```
RESULT 15
US-10-126-052A-263
; Sequence 263, Application US/10126052A
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Murray, Richard
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Lung Cancer, Compositions and
; TITLE OF INVENTION: Methods of Screening for Modulators of Lung Cancer
; FILE REFERENCE: 018501-001530US
; CURRENT APPLICATION NUMBER: US/10/126, 052A
; CURRENT FILING DATE: 2002-04-18
; PRIOR APPLICATION NUMBER: US 60/284, 770
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: US 60/290, 492
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: US 60/339, 245
; PRIOR FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: US 60/350, 666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/334, 370
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/372, 246
; PRIOR FILING DATE: 2002-04-12
; NUMBER OF SEQ ID NOS: 691
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 263
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-126-052A-263
```

```
Query Match          100.0%; Score 26; DB 12; Length 380;
Best Local Similarity 100.0%; Pred No. 4.4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 LYDAK 5
        |||||
Db       121 LYDAK 125
```

Search completed: July 11, 2003, 12:35:11  
Job time : 5.79854 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:54:32 ; Search time 15.738 seconds

(without alignments)  
204.833 Million cell updates/sec

Title: US-10-091-442-8

Perfect score: 28

Sequence: 1 NTEWK 5

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Pending\_Patents\_AA\_Main:\*

1: /cgn2\_6/ptodata/1/paa/US086.COMB.pep:\*

2: /cgn2\_6/ptodata/1/paa/US086.COMB.pep:\*

3: /cgn2\_6/ptodata/1/paa/US087.COMB.pep:\*

4: /cgn2\_6/ptodata/1/paa/US088.COMB.pep:\*

5: /cgn2\_6/ptodata/1/paa/US089.COMB.pep:\*

6: /cgn2\_6/ptodata/1/paa/US090.COMB.pep:\*

7: /cgn2\_6/ptodata/1/paa/US091.COMB.pep:\*

8: /cgn2\_6/ptodata/1/paa/US092.COMB.pep:\*

9: /cgn2\_6/ptodata/1/paa/US093.COMB.pep:\*

10: /cgn2\_6/ptodata/1/paa/US094.COMB.pep:\*

11: /cgn2\_6/ptodata/1/paa/US095.COMB.pep:\*

12: /cgn2\_6/ptodata/1/paa/US096.COMB.pep:\*

13: /cgn2\_6/ptodata/1/paa/US097.COMB.pep:\*

14: /cgn2\_6/ptodata/1/paa/US098.COMB.pep:\*

15: /cgn2\_6/ptodata/1/paa/US099.COMB.pep:\*

16: /cgn2\_6/ptodata/1/paa/US100.COMB.pep:\*

17: /cgn2\_6/ptodata/1/paa/US101.COMB.pep:\*

18: /cgn2\_6/ptodata/1/paa/US102.COMB.pep:\*

19: /cgn2\_6/ptodata/1/paa/US103.COMB.pep:\*

20: /cgn2\_6/ptodata/1/paa/US104.COMB.pep:\*

21: /cgn2\_6/ptodata/1/paa/US105.COMB.pep:\*

22: /cgn2\_6/ptodata/1/paa/US106.COMB.pep:\*

23: /cgn2\_6/ptodata/1/paa/US107.COMB.pep:\*

24: /cgn2\_6/ptodata/1/paa/US108.COMB.pep:\*

25: /cgn2\_6/ptodata/1/paa/US109.COMB.pep:\*

26: /cgn2\_6/ptodata/1/paa/US110.COMB.pep:\*

27: /cgn2\_6/ptodata/1/paa/US111.COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	28	100.0	5	US-08-091-028-8	Sequence 8, Appl1
2	28	100.0	5	US-08-091-028A-8	Sequence 8, Appl1
3	28	100.0	5	US-09-140-719-8	Sequence 8, Appl1
4	28	100.0	5	US-10-091-442-8	Sequence 8, Appl1
5	28	100.0	78	US-60-160-203-4173	Sequence 4173, Ap
6	28	100.0	118	US-60-188-162-4539	Sequence 4539, Ap

7	28	100.0	126	20	US-09-621-976-5429	Sequence 5429, Ap
8	28	100.0	126	17	US-60-147-499-5429	Sequence 5429, Ap
9	28	100.0	145	17	US-09-328-352-6374	Sequence 6374, Ap
10	28	100.0	163	21	US-09-731-872-445	Sequence 445, App
11	28	100.0	163	22	US-09-876-997-445	Sequence 445, App
12	28	100.0	163	27	US-60-169-629-607	Sequence 607, App
13	28	100.0	163	27	US-60-187-470-607	Sequence 607, App
14	28	100.0	181	20	US-09-675-784A-12333	Sequence 12333, A
15	28	100.0	214	16	US-09-252-691-7915	Sequence 7915, Ap
16	28	100.0	214	16	US-09-252-691C-7915	Sequence 7915, Ap
17	28	100.0	278	21	US-09-760-446A-1177	Sequence 1177, Ap
18	28	100.0	294	25	US-10-206-664-1177	Sequence 1177, Ap
19	28	100.0	294	25	US-10-108-260A-4567	Sequence 4567, Ap
20	28	100.0	298	18	US-09-417-507-40692	Sequence 40692, A
21	28	100.0	380	4	US-08-091-028A-34	Sequence 34, Appl
22	28	100.0	380	15	US-09-140-719-34	Sequence 34, Appl
23	28	100.0	380	19	US-09-508-997A-2	Sequence 2, Appl1
24	28	100.0	380	21	US-09-791-537-79163	Sequence 281, App
25	28	100.0	380	23	US-09-936-883A-2	Sequence 2, Appl1
26	28	100.0	380	24	US-10-091-442-34	Sequence 34, Appl
27	28	100.0	380	25	US-10-110-030-2	Sequence 2, Appl1
28	28	100.0	399	24	US-10-029-180-64	Sequence 64, Appl
29	28	100.0	400	16	US-09-248-796-19018	Sequence 19018, A
30	28	100.0	400	27	US-60-096-409-19018	Sequence 19018, A
31	28	100.0	417	21	US-09-760-446A-1893	Sequence 1893, Ap
32	28	100.0	417	26	US-10-206-664-1893	Sequence 1893, Ap
33	28	100.0	420	19	US-09-509-234A-41	Sequence 41, Appl
34	28	100.0	420	19	US-09-509-234C-41	Sequence 41, Appl
35	28	100.0	420	21	US-09-791-537-79163	Sequence 79163, A
36	28	100.0	420	21	US-09-791-537-132920	Sequence 132920, A
37	28	100.0	428	19	US-09-513-996A-37385	Sequence 37385, A
38	28	100.0	428	21	US-09-708-427-10543	Sequence 10543, A
39	28	100.0	441	19	US-09-513-996A-37384	Sequence 37384, A
40	28	100.0	441	19	US-09-708-427-10542	Sequence 10542, A
41	28	100.0	479	19	US-09-513-996A-37383	Sequence 37383, A
42	28	100.0	479	21	US-09-708-427-10541	Sequence 10541, A
43	28	100.0	562	21	US-09-791-537-22455	Sequence 22455, A
44	28	100.0	584	20	US-09-604-957-6	Sequence 6, Appl1
45	28	100.0	584	23	US-09-995-746A-12	Sequence 12, Appl

#### ALIGNMENTS

RESULT 1

US-08-091-028-8

Sequence 8, Application US/08091028

GENERAL INFORMATION:

APPLICANT: TSUJIMOTO, Masafumi

APPLICANT: IMASA, Fuyuki

APPLICANT: TSURUOKA, Nobuo

APPLICANT: NAKAZATO, Hiroshi

APPLICANT: MIURA, Kenju

APPLICANT: ISHIDA, Nobuhiko

APPLICANT: KURIHARA, Tatsuya

APPLICANT: YAMAICHI, Kozo

APPLICANT: YAMAGUCHI, Nozomi

TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR

NUMBER OF SEQUENCES: 33

CORRESPONDENCE ADDRESS:

ADDRESSEE: Burns, Doane, Swecker & Mathis

STREET: George Mason Bldg., Washington & Prince Sts.

CITY: Alexandria

STATE: Virginia

COUNTRY: United States

ZIP: 22313-1404

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/091,028

;; FILING DATE: 14-JUL-1993  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: JP 4-212305  
;; FILING DATE: 17-JUL-1992  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: JP 5-067339  
;; FILING DATE: 04-MAR-1993  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Crane-Feury, Sharon E  
;; REGISTRATION NUMBER: 36,113  
;; REFERENCE/DOCKET NUMBER: 001560-204  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (703) 836-6620  
;; TELEFAX: (703) 836-6620  
;; INFORMATION FOR SEQ ID NO: 8:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 5 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
US-08-091-028-8

Query Match 100.0%; Score 28; DB 4; Length 5;  
Best Local Similarity 100.0%; Pred. No. 4.2e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NYEMK 5  
Db 1 NYEMK 5

## RESULT 2

US-08-091-028A-8  
;; Sequence 8, Application US/08091028A  
;; GENERAL INFORMATION:  
;; APPLICANT: TSUJIMOTO, Masafumi  
;; APPLICANT: IWASA, Fuyuki  
;; APPLICANT: TSURUOKA, Nobuo  
;; APPLICANT: NAKAZATO, Hiroshi  
;; APPLICANT: MURA, Kenju  
;; APPLICANT: ISHIDA, Nobuhiko  
;; APPLICANT: KURIHARA, Tatsuya  
;; APPLICANT: YAMAICHI, Kozo  
;; APPLICANT: YAMAGUCHI, Nozomi  
;; TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
;; NUMBER OF SEQUENCES: 34  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Burns, Doane, Swecker & Mathis  
;; STREET: George Mason Bldg., Washington & Prince Sts.  
;; CITY: Alexandria  
;; STATE: Virginia  
;; COUNTRY: United States  
;; ZIP: 22131-1404  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/091,028A  
;; FILING DATE: 14-JUL-1993  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: JP 4-212305  
;; FILING DATE: 17-JUL-1992  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: JP 5-067339  
;; FILING DATE: 04-MAR-1993  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: REA, TERESA STANEK  
;; REGISTRATION NUMBER: 30,427

;; REFERENCE/DOCKET NUMBER: 001560-204  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (703) 836-6620  
;; TELEFAX: (703) 836-6620  
;; INFORMATION FOR SEQ ID NO: 8:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 5 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
US-08-091-028A-8

Query Match 100.0%; Score 28; DB 4; Length 5;  
Best Local Similarity 100.0%; Pred. No. 4.2e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NYEMK 5  
Db 1 NYEMK 5

## RESULT 3

US-09-140-719-8  
;; Sequence 8, Application US/09140719  
;; GENERAL INFORMATION:  
;; APPLICANT: TSUJIMOTO, Masafumi  
;; APPLICANT: IWASA, Fuyuki  
;; APPLICANT: TSURUOKA, Nobuo  
;; APPLICANT: NAKAZATO, Hiroshi  
;; APPLICANT: MURA, Kenju  
;; APPLICANT: ISHIDA, Nobuhiko  
;; APPLICANT: KURIHARA, Tatsuya  
;; APPLICANT: YAMAICHI, Kozo  
;; APPLICANT: YAMAGUCHI, Nozomi  
;; TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
;; NUMBER OF SEQUENCES: 34  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Burns, Doane, Swecker & Mathis  
;; STREET: P.O. Box 1404  
;; CITY: Alexandria  
;; STATE: Virginia  
;; COUNTRY: United States  
;; ZIP: 22131-1404  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/140,719  
;; FILING DATE: 08-AUG-1998  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/474,661  
;; FILING DATE: 07-JUN-1995  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/091,028  
;; FILING DATE: 14-JUL-1993  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: JP 4-212305  
;; FILING DATE: 17-JUL-1992  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: JP 5-067339  
;; FILING DATE: 04-MAR-1993  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: McGowan, Malcolm K.  
;; REGISTRATION NUMBER: 39,300  
;; REFERENCE/DOCKET NUMBER: 001560-247  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (703) 836-6620  
;; TELEFAX: (703) 836-2021  
;; INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:  
LENGTH: 5 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-140-719-8

Query Match 100.0%; Score 28; DB 15; Length 5;  
Best Local Similarity 100.0%; Pred. No. 4.2e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYEMK 5  
Db 1 NYEMK 5

## RESULT 4

US-10-091-442-8

Sequence 8, Application US/10091442

GENERAL INFORMATION:

APPLICANT: TSUJIMOTO, Masafumi

IMASA, Fuyuki

TSURUDOKA, Nobuo

NAKAZATO, Hiroshi

MURA, Kenju

ISHIDA, Nobuhiko

KURIHARA, Tatsuya

YAMAUCHI, Kozo

YAMAGUCHI, Nozomi

TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR

NUMBER OF SEQUENCES: 34

CORRESPONDENCE ADDRESS:

ADDRESSEE: Burns, Doane, Swecker & Mathis

STREET: P.O. Box 1404

CITY: Alexandria

STATE: Virginia

COUNTRY: United States

ZIP: 22313-1404

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/091,442

FILING DATE: 07-Mar-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/140,719

FILING DATE: 08-AUG-1998

APPLICATION NUMBER: US 08/474,661

FILING DATE: 07-JUN-1995

APPLICATION NUMBER: US 08/091,028

FILING DATE: 14-JUL-1993

APPLICATION NUMBER: JP 4-212305

FILING DATE: 17-JUL-1992

APPLICATION NUMBER: JP 6-067339

FILING DATE: 04-MAR-1993

ATTORNEY/AGENT INFORMATION:

NAME: McGowan, Malcolm K.

REGISTRATION NUMBER: 39,300

REFERENCE/DOCKET NUMBER: 001560-247

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 836-6620

TELEFAX: (703) 836-2021

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 5 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 8:  
US-10-091-442-8

Query Match 100.0%; Score 28; DB 24; Length 5;  
Best Local Similarity 100.0%; Pred. No. 4.2e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYEMK 5  
Db 1 NYEMK 5

## RESULT 5

US-60-160-203-4173

Sequence 4173, Application US/60160203

GENERAL INFORMATION:

APPLICANT: BONAZZI, VIVIAN

TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,

TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS AND

TITLE OF INVENTION: USES THEREOF

FILE REFERENCE: C1000116

CURRENT APPLICATION NUMBER: US/60/160,203

CURRENT FILING DATE: 1999-10-19

NUMBER OF SEQ ID NOS: 6374

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 4173

LENGTH: 78

TYPE: PRT

ORGANISM: HUMAN

US-60-160-203-4173

Query Match 100.0%; Score 28; DB 27; Length 78;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYEMK 5  
Db 37 NYEMK 41

## RESULT 6

US-60-188-162-4539

Sequence 4539, Application US/60188162

GENERAL INFORMATION:

APPLICANT: Beasley, Ellen

TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,

TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND

TITLE OF INVENTION: USES THEREOF

FILE REFERENCE: C1000518

CURRENT APPLICATION NUMBER: US/60/188,162

CURRENT FILING DATE: 2000-05-09

NUMBER OF SEQ ID NOS: 5094

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 4539

LENGTH: 118

TYPE: PRT

ORGANISM: HUMAN

US-60-188-162-4539

Query Match 100.0%; Score 28; DB 27; Length 118;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYEMK 5  
Db 37 NYEMK 41

## RESULT 7

US-09-621-976-5429

Sequence 5429, Application US/09621976

GENERAL INFORMATION:

APPLICANT: Dumas Milne Edwards, J. B.

APPLICANT: Jobert, S.  
APPLICANT: Giordano, J.Y.  
TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
FILE REFERENCE: GENSET.054PR2  
CURRENT APPLICATION NUMBER: US/09/621,976  
CURRENT FILING DATE: 2000-07-21  
NUMBER OF SEQ ID NOS: 19335  
SOFTWARE: Patent.pm  
SEQ ID NO 5429  
LENGTH: 126  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SIGNAL  
LOCATION: -41...-1  
NAME/KEY: UNSURE  
LOCATION: 62  
OTHER INFORMATION: Xaa - Lys,Arg  
US-09-621-976-5429

Query Match 100.0%; Score 28; DB 20; Length 126;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NYEMK 5  
|||||  
DB 122 NYEMK 126

RESULT 8  
US-60-147-499-5429  
Sequence 5429, Application US/60147499  
GENERAL INFORMATION:  
APPLICANT: Dumas Milne Edwards, J.B.  
APPLICANT: Jobert, S.Y.  
TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
FILE REFERENCE: GENSET.054PR2  
CURRENT APPLICATION NUMBER: US/60/147,499  
CURRENT FILING DATE: 1999-08-05  
NUMBER OF SEQ ID NOS: 19335  
SOFTWARE: Patent.pm  
SEQ ID NO 5429  
LENGTH: 126  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SIGNAL  
LOCATION: -41...-1  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: 62  
OTHER INFORMATION: Xaa - Lys,Arg  
US-60-147-499-5429

Query Match 100.0%; Score 28; DB 27; Length 126;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NYEMK 5  
|||||  
DB 122 NYEMK 126

RESULT 9  
US-09-328-352-6374  
Sequence 6374, Application US/09328352  
GENERAL INFORMATION:  
APPLICANT: Gary L. Breton et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
FILE REFERENCE: CTC99-03PA  
CURRENT APPLICATION NUMBER: US/09/328,352

CURRENT FILING DATE: 1999-06-04  
NUMBER OF SEQ ID NOS: 8252  
SEQ ID NO 6374  
LENGTH: 145  
TYPE: PRT  
ORGANISM: Acinetobacter baumannii  
US-09-328-352-6374

Query Match 100.0%; Score 28; DB 17; Length 145;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NYEMK 5  
|||||  
DB 28 NYEMK 32

RESULT 10  
US-09-731-872-445  
Sequence 445, Application US/09731872  
GENERAL INFORMATION:  
APPLICANT: Dumas Milne Edwards, Jean Baptiste  
APPLICANT: Bougueleret, Lydie  
APPLICANT: Jobert, Severin  
TITLE OF INVENTION: FULL-LENGTH HUMAN CDNA ENCODING POTENTIALLY SECRETED PROTEINS  
FILE REFERENCE: 78.US3.REG  
CURRENT APPLICATION NUMBER: US/09/731,872  
CURRENT FILING DATE: 2000-12-07  
PRIOR APPLICATION NUMBER: US 60/169,629  
PRIOR FILING DATE: 1999-12-08  
PRIOR APPLICATION NUMBER: US 60/187,470  
PRIOR FILING DATE: 2000-03-06  
NUMBER OF SEQ ID NOS: 482  
SOFTWARE: Patent.pm  
SEQ ID NO 445  
LENGTH: 163  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-731-872-445

Query Match 100.0%; Score 28; DB 21; Length 163;  
Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NYEMK 5  
|||||  
DB 159 NYEMK 163

RESULT 11  
US-09-876-997-445  
Sequence 445, Application US/09876997  
GENERAL INFORMATION:  
APPLICANT: Dumas Milne Edwards, Jean Baptiste  
APPLICANT: Bougueleret, Lydie  
APPLICANT: Jobert, Severin  
TITLE OF INVENTION: FULL-LENGTH HUMAN CDNA ENCODING POTENTIALLY SECRETED PROTEINS  
FILE REFERENCE: 78.US4.CIP  
CURRENT APPLICATION NUMBER: US/09/876,997  
CURRENT FILING DATE: 2001-06-08  
PRIOR APPLICATION NUMBER: US 09/731,872  
PRIOR FILING DATE: 2000-12-07  
PRIOR APPLICATION NUMBER: US 60/187,470  
PRIOR FILING DATE: 2000-03-06  
PRIOR APPLICATION NUMBER: US 60/169,629  
PRIOR FILING DATE: 1999-12-08  
NUMBER OF SEQ ID NOS: 482  
SOFTWARE: Patent.pm  
SEQ ID NO 445  
LENGTH: 163  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-876-997-445

Query Match 100.0%; Score 28; DB 22; Length 163;  
Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NYEMK 5  
|||||  
Db 159 NYEMK 163

## RESULT 12

US-60-169-629-607  
; Sequence 607, Application US/60169629  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Bougueleret, L.  
; APPLICANT: Jobert, S.  
; TITLE OF INVENTION: CDNAS for Secreted Proteins  
; FILE REFERENCE: GENSET.071PRF  
; CURRENT APPLICATION NUMBER: US/60/169,629  
; CURRENT FILING DATE: 1999-12-08  
; NUMBER OF SEQ ID NOS: 715  
; SOFTWARE: Patent.pm  
; SEQ ID NO 607  
; LENGTH: 163  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: -77...-1  
US-60-169-629-607

Query Match 100.0%; Score 28; DB 27; Length 163;  
Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NYEMK 5  
|||||  
Db 159 NYEMK 163

## RESULT 13

US-60-187-470-607  
; Sequence 607, Application US/60187470  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Bougueleret, L.  
; APPLICANT: Jobert, S.  
; TITLE OF INVENTION: CDNAS for Secreted Proteins  
; FILE REFERENCE: 78.US2.PRO  
; CURRENT APPLICATION NUMBER: US/60/187,470  
; CURRENT FILING DATE: 2000-03-06  
; NUMBER OF SEQ ID NOS: 715  
; SOFTWARE: Patent.pm  
; SEQ ID NO 607  
; LENGTH: 163  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: -77...-1  
US-60-187-470-607

Query Match 100.0%; Score 28; DB 27; Length 163;  
Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NYEMK 5  
|||||  
Db 159 NYEMK 163

## RESULT 14

US-09-675-784A-12333  
; Sequence 12333, Application US/09675784A  
; GENERAL INFORMATION:  
; APPLICANT: HARE, ROBERTA S.  
; APPLICANT: SHAW, KAREN J.  
; APPLICANT: SHIMER JR., GEORGE H.  
; APPLICANT: KESSLER, MARCO  
; APPLICANT: NOLLING, JORCK  
; APPLICANT: ZENG, QIANDONG  
; APPLICANT: GREENE, JONATHAN R.  
; TITLE OF INVENTION: ASPERGILLUS FUMIGATUS NUCLEIC ACIDS AND POLYPEPTIDES,  
; FILE REFERENCE: 2976-4020US1  
; CURRENT APPLICATION NUMBER: US/09/675,784A  
; CURRENT FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/156,338  
; PRIOR FILING DATE: 1999-09-29  
; NUMBER OF SEQ ID NOS: 13925  
; SEQ ID NO 12333  
; LENGTH: 181  
; TYPE: PRT  
; ORGANISM: Aspergillus fumigatus  
US-09-675-784A-12333

Query Match 100.0%; Score 28; DB 20; Length 181;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NYEMK 5  
|||||  
Db 92 NYEMK 96

## RESULT 15

US-09-252-691-7915  
; Sequence 7915, Application US/09252691B  
; GENERAL INFORMATION:  
; APPLICANT: Keith G. Weinstein et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTE  
; FILE REFERENCE: 107196.135  
; CURRENT APPLICATION NUMBER: US/09/252,691B  
; CURRENT FILING DATE: 1999-02-18  
; NUMBER OF SEQ ID NOS: 11324  
; SEQ ID NO 7915  
; LENGTH: 214  
; TYPE: PRT  
; ORGANISM: Enterobacter cloacae  
US-09-252-691-7915

Query Match 100.0%; Score 28; DB 16; Length 214;  
Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NYEMK 5  
|||||  
Db 18 NYEMK 22

Search completed: July 11, 2003, 12:27:38  
Job time: 17.738 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:55:27 ; Search time 4.69854 Seconds  
(without alignments)  
280.699 Million cell updates/sec

Title: US-10-091-442-8

Perfect score: 28  
Sequence: 1 NYEMK 5

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1229476 seqs, 263775074 residues

Total number of hits satisfying chosen parameters: 1229476

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending\_Patents\_AA\_New:\*  
1: /cgn2\_6/ptodata/1/paa/PCT\_NEW\_COMB.pep4:\*  
2: /cgn2\_6/ptodata/1/paa/PCT\_NEW\_COMB.pep4:\*  
3: /cgn2\_6/ptodata/1/paa/US05\_NEW\_COMB.pep4:\*  
4: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pep4:\*  
5: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep4:\*  
6: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep4:\*  
7: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep4:\*  
8: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep4:\*  
9: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep4:\*  
10: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep4:\*  
11: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep4:\*  
12: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep4:\*  
13: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep4:\*  
14: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	28	100.0	116	10	US-09-724-676-95964
2	28	100.0	116	10	US-09-724-676-95964
3	28	100.0	135	12	US-10-424-599-265931
4	28	100.0	145	12	US-10-431-652-6374
5	28	100.0	158	10	US-09-724-676-95968
6	28	100.0	158	10	US-09-724-676-95968
7	28	100.0	175	12	US-10-424-599-254001
8	28	100.0	181	10	US-09-675-784A-12333
9	28	100.0	205	12	US-10-424-599-265929
10	28	100.0	214	12	US-10-417-886-7915
11	28	100.0	380	10	US-09-936-883C-2
12	28	100.0	380	12	US-10-126-052A-263
13	28	100.0	380	14	US-60-443-566-3006
14	28	100.0	380	14	US-60-455-444-4692
15	28	100.0	380	14	US-60-465-241-4692
16	28	100.0	819	14	US-60-443-566-2783
17	28	100.0	819	14	US-60-452-680-17072
18	28	100.0	819	14	US-60-453-135-10384
19	28	100.0	819	14	US-60-453-050-10384

20	28	100.0	819	14	US-60-466-412-10384	Sequence 10384, A
21	28	100.0	847	2	PCT-US02-19669A-498	Sequence 498, App
22	25	89.3	32	12	US-10-019-596A-3	Sequence 3, App11
23	25	89.3	32	12	US-10-019-596A-5	Sequence 5, App11
24	25	89.3	32	12	US-10-019-596A-6	Sequence 6, App11
25	25	89.3	32	12	US-10-019-596A-7	Sequence 7, App11
26	25	89.3	32	12	US-10-019-596A-8	Sequence 8, App11
27	25	89.3	32	12	US-10-019-596A-11	Sequence 11, App1
28	25	89.3	37	10	US-10-019-596A-35	Sequence 35, App1
29	25	89.3	43	10	US-09-950-083B-3405	Sequence 3405, App
30	25	89.3	52	12	US-10-424-599-234765	Sequence 234765, App
31	25	89.3	55	12	US-10-424-599-191002	Sequence 191002, App
32	25	89.3	59	12	US-10-424-599-232746	Sequence 232746, App
33	25	89.3	61	10	US-09-724-676-93757	Sequence 93757, A
34	25	89.3	61	10	US-09-724-676-93757	Sequence 93757, A
35	25	89.3	61	12	US-10-424-599-162592	Sequence 162592, A
36	25	89.3	67	12	US-10-424-599-211936	Sequence 211936, App
37	25	89.3	72	10	US-09-724-676-93759	Sequence 93759, A
38	25	89.3	72	10	US-09-724-676-93759	Sequence 93759, A
39	25	89.3	74	10	US-09-513-999C-7069	Sequence 7069, App
40	25	89.3	74	10	US-09-513-999C-7444	Sequence 7444, App
41	25	89.3	90	10	US-09-513-999C-7444	Sequence 7444, App
42	25	89.3	90	10	US-09-513-999C-7444	Sequence 7444, App
43	25	89.3	101	10	US-09-864-408A-7436	Sequence 7436, App
44	25	89.3	101	12	US-10-437-963-186021	Sequence 186021, App
45	25	89.3	136	12	US-10-437-963-184199	Sequence 184199, App

#### ALIGNMENTS

RESULT 1  
US-09-724-676-95964  
; Sequence 95964, Application US/09724676  
; GENERAL INFORMATION:  
; APPLICANT: Compugen LTD  
; TITLE OF INVENTION: Variants of alternative splicing  
; FILE REFERENCE: 129181.4 Compugen  
; CURRENT APPLICATION NUMBER: US/09/724, 676  
; CURRENT FILING DATE: 2000-11-28  
; NUMBER OF SEQ ID NOS: 97222  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 95964  
; LENGTH: 116  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-09-724-676-95964

Query Match 100.0%; Score 28; DB 10; Length 116;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
LENGTH: 116

QY 1 NYEMK 5  
DB 6 NYEMK 10

RESULT 2  
US-09-724-676A-95964  
; Sequence 95964, Application US/09724676A  
; GENERAL INFORMATION:  
; APPLICANT: Compugen LTD  
; TITLE OF INVENTION: Variants of alternative splicing  
; FILE REFERENCE: 129181.4 Compugen  
; CURRENT APPLICATION NUMBER: US/09/724, 676A  
; CURRENT FILING DATE: 2000-11-28  
; NUMBER OF SEQ ID NOS: 97222  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 95964  
; LENGTH: 116  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-09-724-676A-95964

Query Match 100.0%; Score 28; DB 10; Length 116;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYEMK 5  
11111  
DB 6 NYEMK 10

RESULT 3  
US-10-424-599-265931

; Sequence 265931, Application US/10424599  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 265931  
; LENGTH: 135  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_82156C.1.pep  
US-10-424-599-265931

Query Match 100.0%; Score 28; DB 12; Length 135;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYEMK 5  
11111  
DB 112 NYEMK 116

RESULT 4  
US-10-431-652-6374  
; Sequence 6374, Application US/10431652  
; GENERAL INFORMATION:  
; APPLICANT: Breton, Gary L.  
; APPLICANT: Bush, David  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: PATH03-08  
; CURRENT APPLICATION NUMBER: US/10/431,652  
; CURRENT FILING DATE: 2003-05-06  
; PRIOR APPLICATION NUMBER: US 09/328,352  
; PRIOR FILING DATE: 1999-06-04  
; PRIOR APPLICATION NUMBER: US 60/088,701  
; PRIOR FILING DATE: 1998-06-09  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 6374  
; LENGTH: 145  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-10-431-652-6374

Query Match 100.0%; Score 28; DB 12; Length 145;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYEMK 5  
11111  
DB 28 NYEMK 32

RESULT 5

US-09-724-676-95968  
; Sequence 95968, Application US/09724676

; GENERAL INFORMATION:  
; APPLICANT: Compugen LTD  
; TITLE OF INVENTION: Variants of alternative splicing  
; FILE REFERENCE: 129181.4 Compugen  
; CURRENT APPLICATION NUMBER: US/09/724,676  
; CURRENT FILING DATE: 2000-11-28  
; NUMBER OF SEQ ID NOS: 97222  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 95968  
; LENGTH: 158  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-724-676-95968

Query Match 100.0%; Score 28; DB 10; Length 158;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYEMK 5  
11111  
DB 6 NYEMK 10

RESULT 6  
US-09-724-676A-95968  
; Sequence 95968, Application US/09724676A  
; GENERAL INFORMATION:  
; APPLICANT: Compugen LTD  
; TITLE OF INVENTION: Variants of alternative splicing  
; FILE REFERENCE: 129181.4 Compugen  
; CURRENT APPLICATION NUMBER: US/09/724,676A  
; CURRENT FILING DATE: 2000-11-28  
; NUMBER OF SEQ ID NOS: 97222  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 95968  
; LENGTH: 158  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-724-676A-95968

Query Match 100.0%; Score 28; DB 10; Length 158;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYEMK 5  
11111  
DB 6 NYEMK 10

RESULT 7  
US-10-424-599-254001  
; Sequence 254001, Application US/10424599  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 254001  
; LENGTH: 175  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_71386C.1.pep  
US-10-424-599-254001





Fri Jul 11 15:11:28 2003

us-10-091-442-8.rapn

Page 3

Query Match 100.0%; Score 28; DB 12; Length 175;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYEMK 5  
DB 14 NYEMK 18

RESULT 8

US-09-675-784A-12333  
; Sequence 12333, Application US/09675784A  
; GENERAL INFORMATION:  
; APPLICANT: HARE, ROBERTA S.  
; APPLICANT: SHAW, KAREN J.  
; APPLICANT: SHIMER JR., GEORGE H.  
; APPLICANT: KESSLER, MARCO  
; APPLICANT: NOLLING, JORR  
; APPLICANT: ZENG, QIANDONG  
; APPLICANT: GREENE, JONATHAN R.  
; TITLE OF INVENTION: ASPERGILLUS FUMIGATUS NUCLEIC ACIDS AND POLYPEPTIDES,  
; FILE REFERENCE: 2976-4020U51  
; CURRENT APPLICATION NUMBER: US/09/675,784A  
; CURRENT FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/156,338  
; PRIOR FILING DATE: 1999-09-29  
; NUMBER OF SEQ ID NOS: 13925  
; SEQ ID NO 12333  
; LENGTH: 181  
; TYPE: PRT  
; ORGANISM: Aspergillus fumigatus  
US-09-675-784A-12333

Query Match 100.0%; Score 28; DB 10; Length 181;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYEMK 5  
DB 92 NYEMK 96

RESULT 9

US-10-424-599-265929  
; Sequence 265929, Application US/10424599  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 265929  
; LENGTH: 205  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_82154C.1.pep  
US-10-424-599-265929

Query Match 100.0%; Score 28; DB 12; Length 205;  
Best Local Similarity 100.0%; Pred. No. 48;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYEMK 5  
DB 98 NYEMK 102

RESULT 10

US-10-417-886-7915  
; Sequence 7915, Application US/10417886  
; GENERAL INFORMATION:  
; APPLICANT: Keith G. Weinstein et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER  
; FILE REFERENCE: 107196.135  
; CURRENT APPLICATION NUMBER: US/10/417,886  
; CURRENT FILING DATE: 2003-04-17  
; PRIOR APPLICATION NUMBER: US/09/252,691C  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,145  
; PRIOR FILING DATE: 1998-07-24  
; PRIOR APPLICATION NUMBER: US 60/074,787  
; PRIOR FILING DATE: 1998-02-18  
; NUMBER OF SEQ ID NOS: 11326  
; SEQ ID NO 7915  
; LENGTH: 214  
; TYPE: PRT  
; ORGANISM: Enterobacter cloacae  
US-10-417-886-7915

Query Match 100.0%; Score 28; DB 12; Length 214;  
Best Local Similarity 100.0%; Pred. No. 50;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYEMK 5  
DB 18 NYEMK 22

RESULT 11

US-09-936-883C-2  
; Sequence 2, Application US/09936883C  
; GENERAL INFORMATION:  
; APPLICANT: MIYATA, Toshio  
; TITLE OF INVENTION: A Method for Detecting Megsin Protein and Use  
; FILE REFERENCE: F2-101DIPCT  
; CURRENT APPLICATION NUMBER: US/09/936,883C  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: JP 1999-75305  
; PRIOR FILING DATE: 1999-03-19  
; PRIOR APPLICATION NUMBER: JP 1999-306623  
; PRIOR FILING DATE: 1999-10-28  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 360  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-936-883C-2

Query Match 100.0%; Score 28; DB 10; Length 360;  
Best Local Similarity 100.0%; Pred. No. 81;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NYEMK 5  
DB 284 NYEMK 288

RESULT 12

US-10-126-052A-263  
; Sequence 263, Application US/10126052A  
; GENERAL INFORMATION:  
; APPLICANT: Aziz, Natasha  
; APPLICANT: Murray, Richard  
; APPLICANT: Bos Biotechnology, Inc.  
; TITLE OF INVENTION: Methods of Diagnosis of Lung Cancer, Compositions and  
; TITLE OF INVENTION: Methods of Screening for Modulators of Lung Cancer

```

; FILE REFERENCE: 018501-00130US
; CURRENT APPLICATION NUMBER: US/10/126,052A
; CURRENT FILING DATE: 2002-04-18
; PRIOR APPLICATION NUMBER: US 60/284,770
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: US 60/290,492
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: US 60/339,245
; PRIOR FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/334,370
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/372,246
; PRIOR FILING DATE: 2002-04-12
; NUMBER OF SEQ ID NOS: 691
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 263
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-126-052A-263

```

```

Query Match
Best Local Similarity 100.0%; Score 28; DB 12; Length 380;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 NYEMK 5
    |||||
Db 284 NYEMK 288

```

```

RESULT 13
US-60-443-566-3006
; Sequence 3006, Application US/60443566
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: BEGOVICH, Ann
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001447
; CURRENT APPLICATION NUMBER: US/60/443,566
; CURRENT FILING DATE: 2003-01-30
; NUMBER OF SEQ ID NOS: 25102
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3006
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-443-566-3006

```

```

Query Match
Best Local Similarity 100.0%; Score 28; DB 14; Length 380;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 NYEMK 5
    |||||
Db 284 NYEMK 288

```

```

RESULT 14
US-60-455-444-4692
; Sequence 4692, Application US/60455444
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: BEGOVICH, Ann
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001455
; CURRENT APPLICATION NUMBER: US/60/455,444
; CURRENT FILING DATE: 2003-03-18
; NUMBER OF SEQ ID NOS: 50986
; SOFTWARE: FastSeq for Windows Version 4.0

```

```

; SEQ ID NO 4692
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-455-444-4692

```

```

Query Match
Best Local Similarity 100.0%; Score 28; DB 14; Length 380;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 NYEMK 5
    |||||
Db 284 NYEMK 288

```

```

RESULT 15
US-60-465-241-4692
; Sequence 4692, Application US/60465241
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: BEGOVICH, Ann
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001468
; CURRENT APPLICATION NUMBER: US/60/465,241
; CURRENT FILING DATE: 2003-04-23
; NUMBER OF SEQ ID NOS: 258418
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4692
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-465-241-4692

```

```

Query Match
Best Local Similarity 100.0%; Score 28; DB 14; Length 380;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 NYEMK 5
    |||||
Db 284 NYEMK 288

```

```

Search completed: July 11, 2003, 12:35:12
Job time : 5.79854 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:54:32 ; Search time 31.4761 Seconds

(without alignments)  
204.833 Million cell updates/sec

Title: US-10-091-442-9

Perfect score: 50  
Sequence: 1 AVAMMHQERK 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

Pending\_Patents\_AA\_Main:\*

1: /cgn2\_6/ptodata/1/paa/PCTUS\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/paa/US06\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/paa/US07\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/paa/US08\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/paa/US081\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/paa/US082\_COMB.pep.\*  
7: /cgn2\_6/ptodata/1/paa/US083\_COMB.pep.\*  
8: /cgn2\_6/ptodata/1/paa/US084\_COMB.pep.\*  
9: /cgn2\_6/ptodata/1/paa/US085\_COMB.pep.\*  
10: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep.\*  
11: /cgn2\_6/ptodata/1/paa/US087\_COMB.pep.\*  
12: /cgn2\_6/ptodata/1/paa/US088\_COMB.pep.\*  
13: /cgn2\_6/ptodata/1/paa/US089\_COMB.pep.\*  
14: /cgn2\_6/ptodata/1/paa/US090\_COMB.pep.\*  
15: /cgn2\_6/ptodata/1/paa/US091\_COMB.pep.\*  
16: /cgn2\_6/ptodata/1/paa/US092\_COMB.pep.\*  
17: /cgn2\_6/ptodata/1/paa/US093\_COMB.pep.\*  
18: /cgn2\_6/ptodata/1/paa/US094\_COMB.pep.\*  
19: /cgn2\_6/ptodata/1/paa/US095\_COMB.pep.\*  
20: /cgn2\_6/ptodata/1/paa/US096\_COMB.pep.\*  
21: /cgn2\_6/ptodata/1/paa/US097\_COMB.pep.\*  
22: /cgn2\_6/ptodata/1/paa/US098\_COMB.pep.\*  
23: /cgn2\_6/ptodata/1/paa/US099\_COMB.pep.\*  
24: /cgn2\_6/ptodata/1/paa/US100\_COMB.pep.\*  
25: /cgn2\_6/ptodata/1/paa/US101\_COMB.pep.\*  
26: /cgn2\_6/ptodata/1/paa/US102\_COMB.pep.\*  
27: /cgn2\_6/ptodata/1/paa/US60\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	100.0	10	15	US-09-140-719-9
2	50	100.0	10	24	US-10-091-442-9
3	50	100.0	380	4	US-08-091-028A-34
4	50	100.0	380	15	US-09-140-719-34
5	50	100.0	380	19	US-09-508-997A-2
6	50	100.0	380	21	US-09-791-537-281

7	50	100.0	380	23	US-09-936-883A-2	Sequence 2, Appl1
8	50	100.0	380	25	US-10-091-442-34	Sequence 34, Appl1
9	50	100.0	380	25	US-10-110-030-2	Sequence 2, Appl1
10	44	88.0	10	4	US-08-091-028-9	Sequence 9, Appl1
11	44	88.0	10	4	US-08-091-028A-9	Sequence 4, Appl1
12	41	82.0	380	19	US-09-508-997A-4	Sequence 19, Appl1
13	41	82.0	380	25	US-09-936-883A-19	Sequence 4, Appl1
14	41	82.0	380	25	US-10-110-030-4	Sequence 19, Appl1
15	37	74.0	368	23	US-09-508-997A-6	Sequence 6, Appl1
16	37	74.0	368	23	US-09-936-883A-21	Sequence 21, Appl1
17	37	74.0	368	25	US-10-110-030-6	Sequence 6, Appl1
18	36	72.0	118	1	PCT-US01-08631-41683	Sequence 41683, A
19	36	72.0	336	19	US-09-540-209B-5860	Sequence 5860, Ap
20	35	70.0	212	21	US-09-791-537-79944	Sequence 79944, A
21	35	70.0	576	21	US-09-708-427-16208	Sequence 16208, A
22	35	70.0	578	21	US-09-708-427-2673	Sequence 2673, Ap
23	35	70.0	602	21	US-09-708-427-16207	Sequence 16207, A
24	35	70.0	1373	25	US-60-243-468-1280	Sequence 1280, Ap
25	34	68.0	67	13	US-08-904-468-328	Sequence 328, App
26	34	68.0	67	19	US-09-547-599C-3046	Sequence 3046, Ap
27	34	68.0	76	25	US-10-155-881-19011	Sequence 19011, A
28	34	68.0	82	16	US-09-215-435-380	Sequence 380, App
29	34	68.0	82	22	US-09-834-366-25985	Sequence 25985, A
30	34	68.0	82	22	US-09-856-231A-123	Sequence 123, App
31	34	68.0	82	27	US-60-069-957-248	Sequence 248, App
32	34	68.0	82	27	US-60-197-873-25985	Sequence 25985, A
33	34	68.0	83	1	PCT-US98-12125-155	Sequence 155, App
34	34	68.0	83	16	US-09-209-462B-180	Sequence 180, App
35	34	68.0	83	26	US-10-219-793-160	Sequence 160, App
36	34	68.0	221	1	PCT-US01-08631-41682	Sequence 41682, A
37	34	68.0	980	16	US-09-252-991A-30838	Sequence 30838, A
38	33	66.0	70	27	US-60-177-571-3510	Sequence 3510, Ap
39	33	66.0	70	27	US-60-177-646-2881	Sequence 2881, Ap
40	33	66.0	70	27	US-60-178-311-575	Sequence 575, App
41	33	66.0	108	27	US-60-140-803-509	Sequence 509, App
42	33	66.0	161	27	US-60-142-845-609	Sequence 609, App
43	33	66.0	275	21	US-09-791-537-44885	Sequence 44885, A
44	33	66.0	302	21	US-09-791-537-85844	Sequence 85844, A
45	33	66.0	303	21	US-09-791-537-28004	Sequence 28004, A

## ALIGNMENTS

RESULT 1  
US-09-140-719-9  
Sequence 9, Application US/09140719

GENERAL INFORMATION:

APPLICANT: TSUJIMOTO, Masaaki  
APPLICANT: IMASA, Fuyuki  
APPLICANT: TSUBOYOKA, Nobuo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, Nobuhiko  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAGUCHI, Kozo  
APPLICANT: YAMAGUCHI, Nozomi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719

FILED DATE: 08-AUG-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: MCGOWAN, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-140-719-9

Query Match 100.0%; Score 50; DB 15; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.0027;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AVAMMHQERK 10  
Db 1 AVAMMHQERK 10

## RESULT 2

US-10-091-442-9

Sequence 9, Application US/10091442

GENERAL INFORMATION:

APPLICANT: TSUJIMOTO, Masafumi

IMASA, Fuyuki

TSUROOKA, Nobuo

NAKAZATO, Hiroshi

MURA, Kenju

ISHIDA, Nobuhiko

KURIHARA, Tatsuya

YAMAICHI, Kozo

YAMAGUCHI, Nozomi

TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR

NUMBER OF SEQUENCES: 34

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Burns, Doane, Swecker & Mathis

STREET: P.O. Box 1404

CITY: Alexandria

STATE: Virginia

COUNTRY: United States

ZIP: 22313-1404

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/091,442

FILING DATE: 07-Mar-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/140,719

FILING DATE: 08-AUG-1998

APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: MCGOWAN, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 9:  
US-10-091-442-9

Query Match 100.0%; Score 50; DB 24; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.0027;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AVAMMHQERK 10  
Db 1 AVAMMHQERK 10

## RESULT 3

US-08-091-028A-34

Sequence 34, Application US/08091028A

GENERAL INFORMATION:

APPLICANT: TSUJIMOTO, Masafumi

IMASA, Fuyuki

TSUROOKA, Nobuo

NAKAZATO, Hiroshi

MURA, Kenju

ISHIDA, Nobuhiko

KURIHARA, Tatsuya

YAMAICHI, Kozo

YAMAGUCHI, Nozomi

TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR

NUMBER OF SEQUENCES: 34

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Burns, Doane, Swecker & Mathis

STREET: George Mason Bldg., Washington & Prince Sts.

CITY: Alexandria

STATE: Virginia

COUNTRY: United States

ZIP: 22313-1404

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/091,028A

FILING DATE: 14-JUL-1993

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 4-212305

FILING DATE: 17-JUL-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 5-067339

FILING DATE: 04-MAR-1993

ATTORNEY/AGENT INFORMATION:

NAME: REA, TERESA STANER

REGISTRATION NUMBER: 30,427  
REFERENCE/DOCKET NUMBER: 001560-204  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-6620  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-091-028A-34

Query Match  
Best Local Similarity 100.0%; Score 50; DB 4; Length 380;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AVAMMHOERK 10  
Db 204 AVAMMHOERK 213

RESULT 4  
US-09-140-719-34  
Sequence 34, Application US/09140719  
GENERAL INFORMATION:  
APPLICANT: TSUTSUMOTO, Masaaki  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSURUOKA, Nobuo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, Nobuhiko  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAGUCHI, Kozi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 34:

SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-140-719-34

Query Match  
Best Local Similarity 100.0%; Score 50; DB 15; Length 380;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AVAMMHOERK 10  
Db 204 AVAMMHOERK 213

RESULT 5  
US-09-508-997A-2  
Sequence 2, Application US/09508997A  
GENERAL INFORMATION:  
APPLICANT: MIYATA, Toshio  
TITLE OF INVENTION: Megsin Protein  
FILE REFERENCE: SHIM-004  
CURRENT APPLICATION NUMBER: US/09/508,997A  
CURRENT FILING DATE: 2000-06-27  
PRIOR APPLICATION NUMBER: JP 09-275302  
PRIOR FILING DATE: 1997-09-22  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 2  
LENGTH: 380  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-508-997A-2

Query Match  
Best Local Similarity 100.0%; Score 50; DB 19; Length 380;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AVAMMHOERK 10  
Db 204 AVAMMHOERK 213

RESULT 6  
US-09-791-537-281  
Sequence 281, Application US/09791537  
GENERAL INFORMATION:  
APPLICANT: Biocomix, Inc.  
APPLICANT: Debe, Derek  
TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME  
FILE REFERENCE: 261/210  
CURRENT APPLICATION NUMBER: US/09/791,537  
CURRENT FILING DATE: 2001-02-22  
NUMBER OF SEQ ID NOS: 153055  
SOFTWARE: Patent In version 3.0  
SEQ ID NO 281  
LENGTH: 380  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-791-537-281

Query Match  
Best Local Similarity 100.0%; Score 50; DB 21; Length 380;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AVAMMHOERK 10  
Db 204 AVAMMHOERK 213

RESULT 7  
US-09-936-883A-2  
Sequence 2, Application US/09936883A  
GENERAL INFORMATION:  
APPLICANT: MIYATA, Toshio  
TITLE OF INVENTION: A Method for Detecting Megsin Protein and Use  
FILE REFERENCE: F2-101DPICT  
CURRENT APPLICATION NUMBER: US/09/936, 883A  
PRIOR FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: JP 1999-75305  
PRIOR FILING DATE: 1999-03-19  
PRIOR APPLICATION NUMBER: JP 1999-306623  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 2  
LENGTH: 380  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-936-883A-2

Query Match 100.0%; Score 50; DB 23; Length 380;  
Best Local Similarity 100.0%; Pred. No. 0.21;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AVAMHQERK 10  
DB 204 AVAMHQERK 213

RESULT 8  
US-10-091-442-34  
Sequence 34, Application US/10091442  
GENERAL INFORMATION:  
APPLICANT: TSUTIMOTO, Masafumi  
IMASA, Fuyuki  
TSURUOKA, NODUO  
NAKAZATO, Hiroshi  
MURA, Kenju  
ISHIDA, Nobuhiko  
KURIHARA, Tatsuya  
YAMAICHI, Kozo  
YAMAGUCHI, Nozomi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/091,442  
FILING DATE: 07-May-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993

ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 34:  
US-10-091-442-34

Query Match 100.0%; Score 50; DB 24; Length 380;  
Best Local Similarity 100.0%; Pred. No. 0.21;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AVAMHQERK 10  
DB 204 AVAMHQERK 213

RESULT 9  
US-10-110-030-2  
Sequence 2, Application US/10110030  
GENERAL INFORMATION:  
APPLICANT: Miyata, Toshio  
TITLE OF INVENTION: ANIMAL MODEL FOR MESSENGIAL PROLIFERATIVE  
FILE REFERENCE: SHIM015  
CURRENT APPLICATION NUMBER: US/10/110,030  
PRIOR FILING DATE: 2002-04-03  
PRIOR APPLICATION NUMBER: 11/285736  
PRIOR FILING DATE: 1999-10-06  
NUMBER OF SEQ ID NOS: 15  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 380  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-110-030-2

Query Match 100.0%; Score 50; DB 25; Length 380;  
Best Local Similarity 100.0%; Pred. No. 0.21;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AVAMHQERK 10  
DB 204 AVAMHQERK 213

RESULT 10  
US-08-091-028-9  
Sequence 9, Application US/08091028  
GENERAL INFORMATION:  
APPLICANT: TSUTIMOTO, Masafumi  
IMASA, Fuyuki  
TSURUOKA, NODUO  
NAKAZATO, Hiroshi  
MURA, Kenju  
ISHIDA, Nobuhiko  
KURIHARA, Tatsuya  
YAMAICHI, Kozo  
YAMAGUCHI, Nozomi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria

STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/091,028  
FILING DATE: 14-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA: JP 5-067339  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Crane-Feuzy, Sharon E  
REGISTRATION NUMBER: 36,113  
REFERENCE/DOCKET NUMBER: 001560-204  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-6620  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-091-028-9

Query Match 88.0%; Score 44; DB 4; Length 10;  
Best Local Similarity 90.0%; Pred. No. 0.048;  
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AVAMHQRK 10  
DB 1 AVAMHQRK 10

RESULT 11  
US-08-091-028A-9  
Sequence 9, Application US/08091028A  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IMASA, Fuyuki  
APPLICANT: TSUBUOKA, Nobuo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, Nobuhito  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, Nozomi  
TITLE OF INVENTION: MEGAKARIOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/091,028A  
FILING DATE: 14-JUL-1993

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA: JP 5-067339  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: REA, TERESA SPANER  
REGISTRATION NUMBER: 30,427  
REFERENCE/DOCKET NUMBER: 001560-204  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-6620  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-091-028A-9

Query Match 88.0%; Score 44; DB 4; Length 10;  
Best Local Similarity 90.0%; Pred. No. 0.048;  
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AVAMHQRK 10  
DB 1 AVAMHQRK 10

RESULT 12  
US-09-508-997A-4  
Sequence 4, Application US/09508997A  
GENERAL INFORMATION:  
APPLICANT: MIYATA, Toshio  
TITLE OF INVENTION: Megsin Protein  
FILE REFERENCE: SHIM-004  
CURRENT APPLICATION NUMBER: US/09/508,997A  
CURRENT FILING DATE: 2000-06-27  
PRIOR APPLICATION NUMBER: JP 09-275302  
PRIOR FILING DATE: 1997-09-22  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 4  
LENGTH: 380  
TYPE: PRT  
ORGANISM: Rattus rattus  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: 51, 94  
OTHER INFORMATION: Xaa is unknown.  
US-09-508-997A-4

Query Match 82.0%; Score 41; DB 19; Length 380;  
Best Local Similarity 80.0%; Pred. No. 16;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 AVAMHQRK 10  
DB 204 AVAMHQRK 213

RESULT 13  
US-09-936-883A-19  
Sequence 19, Application US/09936883A  
GENERAL INFORMATION:  
APPLICANT: MIYATA, Toshio  
TITLE OF INVENTION: A Method for Detecting Megsin Protein and Use  
FILE REFERENCE: F2-101DIPCT  
CURRENT APPLICATION NUMBER: US/09/936,883A

```
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: JP 1999-75305
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: JP 1999-306623
; PRIOR FILING DATE: 1999-10-28
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 19
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Rattus norvegicus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 51, 94
; OTHER INFORMATION: Xaa is unknown.
; US-09-936-883A-19
```

```
Query Match      82.0%; Score 41; DB 23; Length 380;
Best Local Similarity 80.0%; Pred. No. 16;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 AVAMMHQERR 10
        |||||||
DB      204 AVNMHQRERR 213
```

```
RESULT 14
US-10-110-030-4
; Sequence 4, Application US/10110030
; GENERAL INFORMATION:
; APPLICANT: Miyata, Toshio
; TITLE OF INVENTION: ANIMAL MODEL FOR MESENTERIAL PROLIFERATIVE
; FILE REFERENCE: SHIM015
; CURRENT APPLICATION NUMBER: US/10/110,030
; PRIOR FILING DATE: 2002-04-03
; PRIOR APPLICATION NUMBER: 11/285736
; PRIOR FILING DATE: 1999-10-06
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Rattus norvegicus
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 51, 94
; OTHER INFORMATION: Xaa = Any Amino Acid
; US-10-110-030-4
```

```
Query Match      82.0%; Score 41; DB 25; Length 380;
Best Local Similarity 80.0%; Pred. No. 16;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 AVAMMHQERR 10
        |||||||
DB      204 AVNMHQRERR 213
```

```
RESULT 15
US-09-508-997A-6
; Sequence 6, Application US/09508997A
; GENERAL INFORMATION:
; APPLICANT: MIYATA, Toshio
; TITLE OF INVENTION: Mesgin Protein
; FILE REFERENCE: SHIM-004
; CURRENT APPLICATION NUMBER: US/09/508,997A
; PRIOR FILING DATE: 2000-06-27
; PRIOR APPLICATION NUMBER: JP 09-275302
; PRIOR FILING DATE: 1997-09-22
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 6
```

```
; LENGTH: 368
; TYPE: PRT
; ORGANISM: Mus musculus
; US-09-508-997A-6
```

```
Query Match      74.0%; Score 37; DB 19; Length 368;
Best Local Similarity 77.8%; Pred. No. 1e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      2 VAMMHQERR 10
        |||||||
DB      193 VNMHQRERR 201
```

```
Search completed: July 11, 2003, 12:27:40
Job time : 33.4761 secs
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: July 11, 2003, 11:55:27 ; Search time 9.39709 Seconds  
(without alignments)  
280.699 Million cell updates/sec

Title: US-10-091-442-9  
Perfect score: 50  
Sequence: 1 AVAMHQRK 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1229476 seqs, 263775074 residues

Total number of hits satisfying chosen parameters: 1229476

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

## Database :

Pending\_Patents\_AA\_New.\*  
1: /cgn2\_6/ptodata/1/paa/PCT\_NEW\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/paa/PCT\_NEW\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep.\*  
7: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep.\*  
8: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep.\*  
9: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep.\*  
10: /cgn2\_6/ptodata/1/paa/US05\_NEW\_COMB.pep.\*  
11: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep.\*  
12: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep.\*  
13: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep.\*  
14: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	50	100.0	380	10	US-09-936-883C-2
2	50	100.0	380	12	US-10-126-052A-263
3	50	100.0	380	14	US-60-443-566-3006
4	50	100.0	380	14	US-60-455-444-4692
5	50	100.0	380	14	US-60-465-241-4692
6	41	82.0	380	10	US-09-936-883C-19
7	40	80.0	131	10	US-09-864-408A-6424
8	37	74.0	368	10	US-09-936-883C-21
9	34	68.0	48	14	US-60-443-566-4420
10	34	68.0	69	14	US-60-443-566-4419
11	34	68.0	82	10	US-09-724-676-86949
12	34	68.0	82	10	US-09-724-676A-86949
13	34	68.0	82	14	US-60-443-566-4418
14	34	68.0	83	12	US-10-219-051B-5879
15	34	68.0	83	12	US-10-219-051B-5882
16	34	68.0	698	12	US-10-437-963-172615
17	34	68.0	980	12	US-10-366-683-30838
18	34	68.0	980	12	US-10-419-128-30838
19	33	66.0	161	12	US-10-437-963-160324

20	33	66.0	337	12	US-10-289-762-411	Sequence 411, App
21	33	66.0	453	12	US-10-437-963-104627	Sequence 104627, A
22	33	66.0	466	12	US-10-366-683-17916	Sequence 17916, A
23	33	66.0	466	12	US-10-419-128-17916	Sequence 17916, A
24	33	66.0	560	10	US-09-949-016-10786	Sequence 10786, A
25	33	66.0	560	12	US-10-087-192-1020	Sequence 22017, Ap
26	33	66.0	646	12	US-10-227-425-2207	Sequence 22017, Ap
27	33	66.0	655	14	US-60-452-680-24115	Sequence 24115, A
28	33	66.0	700	12	US-10-276-781-1352	Sequence 1352, Ap
29	33	66.0	700	12	US-10-366-547-63	Sequence 63, Appl
30	33	66.0	766	12	US-10-219-051B-1638	Sequence 12638, A
31	32	64.0	96	12	US-10-437-963-179299	Sequence 179299, A
32	32	64.0	136	12	US-10-264-237-2376	Sequence 2376, Ap
33	32	64.0	149	12	US-10-437-963-154295	Sequence 154295, A
34	32	64.0	220	12	US-10-424-599-210174	Sequence 210174, A
35	32	64.0	221	12	US-10-425-114-38163	Sequence 38163, A
36	32	64.0	225	12	US-10-266-887-19	Sequence 19, Appl
37	32	64.0	279	10	US-09-858-332B-14	Sequence 14, Appl
38	32	64.0	279	10	US-09-858-332B-14	Sequence 14, Appl
39	32	64.0	279	10	US-09-858-332C-14	Sequence 14, Appl
40	32	64.0	298	12	US-10-282-122A-62871	Sequence 62871, A
41	32	64.0	298	12	US-10-282-122A-64333	Sequence 64333, A
42	32	64.0	304	12	US-10-282-122A-62036	Sequence 62036, A
43	32	64.0	493	12	US-10-366-683-28024	Sequence 28024, A
44	32	64.0	493	12	US-10-419-128-28024	Sequence 28024, A
45	32	64.0	496	12	US-10-282-122A-76573	Sequence 76573, A

## ALIGNMENTS

RESULT 1  
US-09-936-883C-2  
Sequence 2, Appl  
GENERAL INFORMATION:  
APPLICANT: MIYATA, Toshio  
TITLE OF INVENTION: A Method for Detecting Megsin Protein and Use  
FILE REFERENCE: P2-101DIPCT  
CURRENT APPLICATION NUMBER: US/09/936, 883C  
CURRENT FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: JP 1999-75305  
PRIOR FILING DATE: 1999-03-19  
PRIOR APPLICATION NUMBER: JP 1999-306623  
PRIOR FILING DATE: 1999-10-28  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 380  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-936-883C-2

Query Match 100.0%; Score 50; DB 10; Length 380;  
Best Local Similarity 100.0%; Pred. No. 0.12;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AVAMHQRK 10  
|||||||  
Db 204 AVAMHQRK 213

RESULT 2  
US-10-126-052A-263  
Sequence 263, Appl  
GENERAL INFORMATION:  
APPLICANT: Aziz, Natasha  
APPLICANT: Murray, Richard  
APPLICANT: Bos Biotechnology, Inc.  
TITLE OF INVENTION: Methods of Diagnosis of Lung Cancer, Compositions and  
FILE REFERENCE: 018501-001530US  
CURRENT APPLICATION NUMBER: US/10/126, 052A

```
; CURRENT FILING DATE: 2002-04-18
; PRIOR APPLICATION NUMBER: US 60/284,770
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: US 60/290,492
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: US 60/339,245
; PRIOR FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/334,370
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/372,246
; PRIOR FILING DATE: 2002-04-12
; NUMBER OF SEQ ID NOS: 691
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 263
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-126-052A-263
```

```
Query Match
Best Local Similarity 100.0%; Score 50; DB 12; Length 380;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 AVAMMHQERK 10
Db 204 AVAMMHQERK 213
```

```
RESULT 3
US-60-443-566-3006
; Sequence 3006, Application US/60443566
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: BEGOVICH, Ann
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001447
; CURRENT APPLICATION NUMBER: US/60/443,566
; CURRENT FILING DATE: 2003-01-30
; NUMBER OF SEQ ID NOS: 25102
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3006
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-443-566-3006
```

```
Query Match
Best Local Similarity 100.0%; Score 50; DB 14; Length 380;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 AVAMMHQERK 10
Db 204 AVAMMHQERK 213
```

```
RESULT 4
US-60-455-444-4692
; Sequence 4692, Application US/60455444
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: BEGOVICH, Ann
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001455
; CURRENT APPLICATION NUMBER: US/60/455,444
; CURRENT FILING DATE: 2003-03-18
; NUMBER OF SEQ ID NOS: 50986
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4692
; LENGTH: 380
```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-455-444-4692
```

```
Query Match
Best Local Similarity 100.0%; Score 50; DB 14; Length 380;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 AVAMMHQERK 10
Db 204 AVAMMHQERK 213
```

```
RESULT 5
US-60-465-241-4692
; Sequence 4692, Application US/60465241
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: BEGOVICH, Ann
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001468
; CURRENT APPLICATION NUMBER: US/60/465,241
; CURRENT FILING DATE: 2003-04-23
; NUMBER OF SEQ ID NOS: 258418
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4692
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-465-241-4692
```

```
Query Match
Best Local Similarity 100.0%; Score 50; DB 14; Length 380;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 AVAMMHQERK 10
Db 204 AVAMMHQERK 213
```

```
RESULT 6
US-09-936-883C-19
; Sequence 19, Application US/09936883C
; GENERAL INFORMATION:
; APPLICANT: MIYATA, Toshio
; TITLE OF INVENTION: A Method for Detecting Megsin Protein and Use
; FILE REFERENCE: F2-101DIP1PCT
; CURRENT APPLICATION NUMBER: US/09/936,883C
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: JP 1999-75305
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: JP 1999-306623
; PRIOR FILING DATE: 1999-10-28
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 19
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Rattus norvegicus
; NAME/KEY: misc.feature
; LOCATION: 51, 94
; OTHER INFORMATION: Xaa is unknown.
US-09-936-883C-19
```

```
Query Match
Best Local Similarity 82.0%; Score 41; DB 10; Length 380;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 1 AVAMMHQERK 10
Db 204 AVAMMHQERK 213
```



```

US-09-724-676A-86949
; Sequence 86949, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09724, 676A
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 86949
; LENGTH: 82
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-86949

```

```

Query Match      68.0%; Score 34; DB 10; Length 82;
Best Local Similarity 55.6%; Pred. No. 33;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 AVAMMHQER 9
      1:1:1:1:1:1
Db      18 AIAVLHEER 26

```

```

RESULT 13
US-60-443-566-4418
; Sequence 4418, Application US/60443566
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: BEGOVICH, Ann
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001447
; FILE REFERENCE: RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
; CURRENT APPLICATION NUMBER: US/60/443,566
; CURRENT FILING DATE: 2003-01-30
; NUMBER OF SEQ ID NOS: 25102
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4418
; LENGTH: 82
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-443-566-4418

```

```

Query Match      68.0%; Score 34; DB 14; Length 82;
Best Local Similarity 55.6%; Pred. No. 33;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 AVAMMHQER 9
      1:1:1:1:1:1
Db      18 AIAVLHEER 26

```

```

RESULT 14
US-10-219-051B-5879
; Sequence 5879, Application US/10219051B
; GENERAL INFORMATION:
; APPLICANT: The General Hospital Corporation doing business as Massachusetts General
; APPLICANT: Hospital / Bayer AG
; TITLE OF INVENTION: Nucleotide sequences involved in pain
; FILE REFERENCE: Lea 35693 Foreign Counties
; CURRENT APPLICATION NUMBER: US/10/219, 051B
; CURRENT FILING DATE: 2003-05-09
; PRIOR APPLICATION NUMBER: US 60/312,147
; PRIOR FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: US 60/346,382
; PRIOR FILING DATE: 2001-11-01
; PRIOR APPLICATION NUMBER: US 60/333,347
; PRIOR FILING DATE: 2001-11-26
; NUMBER OF SEQ ID NOS: 14715
; SOFTWARE: Perl script
; SEQ ID NO 5879
; LENGTH: 83

```

```

; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(83)
; OTHER INFORMATION: Xaa=Unknown amino acid
US-10-219-051B-5879

```

```

Query Match      68.0%; Score 34; DB 12; Length 83;
Best Local Similarity 55.6%; Pred. No. 34;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 AVAMMHQER 9
      1:1:1:1:1:1
Db      18 AIAVLHEER 26

```

```

RESULT 15
US-10-219-051B-5882
; Sequence 5882, Application US/10219051B
; GENERAL INFORMATION:
; APPLICANT: The General Hospital Corporation doing business as Massachusetts General
; APPLICANT: Hospital / Bayer AG
; TITLE OF INVENTION: Nucleotide sequences involved in pain
; FILE REFERENCE: Lea 35693 Foreign Counties
; CURRENT APPLICATION NUMBER: US/10/219, 051B
; CURRENT FILING DATE: 2003-05-09
; PRIOR APPLICATION NUMBER: US 60/312,147
; PRIOR FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: US 60/346,382
; PRIOR FILING DATE: 2001-11-01
; PRIOR APPLICATION NUMBER: US 60/333,347
; PRIOR FILING DATE: 2001-11-26
; NUMBER OF SEQ ID NOS: 14715
; SOFTWARE: Perl script
; SEQ ID NO 5882
; LENGTH: 83
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(83)
; OTHER INFORMATION: Xaa=Unknown amino acid
US-10-219-051B-5882

```

```

Query Match      68.0%; Score 34; DB 12; Length 83;
Best Local Similarity 55.6%; Pred. No. 34;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 AVAMMHQER 9
      1:1:1:1:1:1
Db      18 AIAVLHEER 26

```

```

Search completed: July 11, 2003, 12:35:13
Job time : 10.4971 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:54:32 ; Search time 1196.09 Seconds  
(Without alignments)  
204.833 Million cell updates/sec

Title: US-10-091-442-34

Perfect score: 1959

Sequence: 1 MASLAANAFECNLFREMD.....FLFVIRKDIILFGKVS CP 380

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Pending-Patents\_AA\_Main:\*

1: /cgn2\_6/ptodata/1/paa/PCRTUS\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/paa/US06\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/paa/US07\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/paa/US080\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/paa/US081\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/paa/US082\_COMB.pep.\*  
7: /cgn2\_6/ptodata/1/paa/US083\_COMB.pep.\*  
8: /cgn2\_6/ptodata/1/paa/US084\_COMB.pep.\*  
9: /cgn2\_6/ptodata/1/paa/US085\_COMB.pep.\*  
10: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep.\*  
11: /cgn2\_6/ptodata/1/paa/US087\_COMB.pep.\*  
12: /cgn2\_6/ptodata/1/paa/US088\_COMB.pep.\*  
13: /cgn2\_6/ptodata/1/paa/US089\_COMB.pep.\*  
14: /cgn2\_6/ptodata/1/paa/US090\_COMB.pep.\*  
15: /cgn2\_6/ptodata/1/paa/US091\_COMB.pep.\*  
16: /cgn2\_6/ptodata/1/paa/US092\_COMB.pep.\*  
17: /cgn2\_6/ptodata/1/paa/US093\_COMB.pep.\*  
18: /cgn2\_6/ptodata/1/paa/US094\_COMB.pep.\*  
19: /cgn2\_6/ptodata/1/paa/US095\_COMB.pep.\*  
20: /cgn2\_6/ptodata/1/paa/US096\_COMB.pep.\*  
21: /cgn2\_6/ptodata/1/paa/US097\_COMB.pep.\*  
22: /cgn2\_6/ptodata/1/paa/US098\_COMB.pep.\*  
23: /cgn2\_6/ptodata/1/paa/US099\_COMB.pep.\*  
24: /cgn2\_6/ptodata/1/paa/US100\_COMB.pep.\*  
25: /cgn2\_6/ptodata/1/paa/US101\_COMB.pep.\*  
26: /cgn2\_6/ptodata/1/paa/US102\_COMB.pep.\*  
27: /cgn2\_6/ptodata/1/paa/US60\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1959	100.0	380	US-08-091-028A-34
2	1959	100.0	380	US-09-140-719-34
3	1959	100.0	380	US-09-508-997A-2
4	1959	100.0	380	US-09-791-537-281
5	1959	100.0	380	US-09-936-883A-2
6	1959	100.0	380	US-10-091-442-34

7	1959	100.0	380	25	US-10-110-030-2	Sequence 2, Appl 1
8	1477	75.4	380	19 <td>US-09-508-997A-4</td> <td>Sequence 19, Appl 1</td>	US-09-508-997A-4	Sequence 19, Appl 1
9	1477	75.4	380	23 <td>US-09-936-883A-19</td> <td>Sequence 4, Appl 1</td>	US-09-936-883A-19	Sequence 4, Appl 1
10	1477	75.4	380	25 <td>US-10-110-030-4</td> <td>Sequence 6, Appl 1</td>	US-10-110-030-4	Sequence 6, Appl 1
11	1458	74.4	368	19 <td>US-09-508-997A-6</td> <td>Sequence 21, Appl 1</td>	US-09-508-997A-6	Sequence 21, Appl 1
12	1458	74.4	368	23 <td>US-09-936-883A-21</td> <td>Sequence 6, Appl 1</td>	US-09-936-883A-21	Sequence 6, Appl 1
13	1458	74.4	368	25 <td>US-10-110-030-6</td> <td>Sequence 946, App</td>	US-10-110-030-6	Sequence 946, App
14	796.5	40.7	379	16 <td>US-09-852-911-946</td> <td>Sequence 946, App</td>	US-09-852-911-946	Sequence 946, App
15	796.5	40.7	379	22 <td>US-09-895-814-946</td> <td>Sequence 946, App</td>	US-09-895-814-946	Sequence 946, App
16	796.5	40.7	379	24 <td>US-10-006-920-946</td> <td>Sequence 946, App</td>	US-10-006-920-946	Sequence 946, App
17	796.5	40.7	379	24 <td>US-10-012-896-946</td> <td>Sequence 946, App</td>	US-10-012-896-946	Sequence 946, App
18	796.5	40.7	379	25 <td>US-10-144-678A-946</td> <td>Sequence 2, Appl 1</td>	US-10-144-678A-946	Sequence 2, Appl 1
19	771.5	39.4	395	22 <td>US-09-823-187-2</td> <td>Sequence 59, Appl 1</td>	US-09-823-187-2	Sequence 59, Appl 1
20	769	39.3	394	1 <td>PCT-US01-13360A-59</td> <td>Sequence 82, Appl 1</td>	PCT-US01-13360A-59	Sequence 82, Appl 1
21	769	39.3	394	1 <td>PCT-US01-13360A-59</td> <td>Sequence 83, Appl 1</td>	PCT-US01-13360A-59	Sequence 83, Appl 1
22	763.5	39.0	405	24 <td>US-10-023-634-82</td> <td>Sequence 2, Appl 1</td>	US-10-023-634-82	Sequence 2, Appl 1
23	763.5	39.0	405	27 <td>US-60-341-362-12</td> <td>Sequence 2, Appl 1</td>	US-60-341-362-12	Sequence 2, Appl 1
24	763.5	39.0	423	24 <td>US-10-023-634-83</td> <td>Sequence 12, Appl 1</td>	US-10-023-634-83	Sequence 12, Appl 1
25	763.5	39.0	423	24 <td>US-10-023-634-83</td> <td>Sequence 12, Appl 1</td>	US-10-023-634-83	Sequence 12, Appl 1
26	754.5	38.5	425	1 <td>PCT-US02-07215-2</td> <td>Sequence 2, Appl 1</td>	PCT-US02-07215-2	Sequence 2, Appl 1
27	754.5	38.5	425	24 <td>US-10-024-427-2</td> <td>Sequence 2, Appl 1</td>	US-10-024-427-2	Sequence 2, Appl 1
28	754.5	38.5	425	24 <td>US-10-094-944-2</td> <td>Sequence 2, Appl 1</td>	US-10-094-944-2	Sequence 2, Appl 1
29	753.5	38.5	425	27 <td>US-60-341-362-4</td> <td>Sequence 12, Appl 1</td>	US-60-341-362-4	Sequence 12, Appl 1
30	752	38.4	390	1 <td>PCT-US02-07215-12</td> <td>Sequence 2, Appl 1</td>	PCT-US02-07215-12	Sequence 2, Appl 1
31	752	38.4	390	4 <td>US-08-099-259-2</td> <td>Sequence 2, Appl 1</td>	US-08-099-259-2	Sequence 2, Appl 1
32	752	38.4	390	9 <td>US-08-568-147A-2</td> <td>Sequence 9, Appl 1</td>	US-08-568-147A-2	Sequence 9, Appl 1
33	752	38.4	390	20 <td>US-09-653-464B-9</td> <td>Sequence 3386, A</td>	US-09-653-464B-9	Sequence 3386, A
34	752	38.4	390	21 <td>US-09-791-537-3386</td> <td>Sequence 76123, A</td>	US-09-791-537-3386	Sequence 76123, A
35	752	38.4	390	21 <td>US-09-791-537-76123</td> <td>Sequence 121576, A</td>	US-09-791-537-76123	Sequence 121576, A
36	752	38.4	390	21 <td>US-09-791-537-76123</td> <td>Sequence 29, Appl 1</td>	US-09-791-537-76123	Sequence 29, Appl 1
37	752	38.4	390	22 <td>US-09-823-187-29</td> <td>Sequence 30, Appl 1</td>	US-09-823-187-29	Sequence 30, Appl 1
38	752	38.4	390	22 <td>US-09-823-187-30</td> <td>Sequence 65, Appl 1</td>	US-09-823-187-30	Sequence 65, Appl 1
39	752	38.4	390	24 <td>US-10-037-417-65</td> <td>Sequence 12, Appl 1</td>	US-10-037-417-65	Sequence 12, Appl 1
40	752	38.4	390	24 <td>US-10-094-944-12</td> <td>Sequence 106, App</td>	US-10-094-944-12	Sequence 106, App
41	752	38.4	413	21 <td>US-09-760-492-106</td> <td>Sequence 2, Appl 1</td>	US-09-760-492-106	Sequence 2, Appl 1
42	752	38.4	413	25 <td>US-10-162-749-106</td> <td>Sequence 13, Appl 1</td>	US-10-162-749-106	Sequence 13, Appl 1
43	750.5	38.3	425	25 <td>US-10-113-113-2</td> <td>Sequence 10, Appl 1</td>	US-10-113-113-2	Sequence 10, Appl 1
44	748	38.2	390	1 <td>PCT-US02-07215-13</td> <td></td>	PCT-US02-07215-13	
45	748	38.2	390	20 <td>US-09-653-464B-10</td> <td></td>	US-09-653-464B-10	

## ALIGNMENTS

RESULT 1

US-08-091-028A-34

Sequence 34, Application US/08091028A

GENERAL INFORMATION:

APPLICANT: TSUJIMOTO, Masafumi

APPLICANT: IMASA, Fuyuki

APPLICANT: TSURUOKA, Nobuo

APPLICANT: NAKAZATO, Hiroshi

APPLICANT: MIURA, Kenju

APPLICANT: ISHIDA, Nobuhiko

APPLICANT: KURIHARA, Tatsuya

APPLICANT: YAMAGUCHI, Kozo

APPLICANT: YAMAGUCHI, Nozomi

TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR

NUMBER OF SEQUENCES: 34

CORRESPONDENCE ADDRESS:

ADDRESS: Burns, Doane, Swecker & Mathis

STREET: George Mason Bldg., Washington & Prince Sts.

CITY: Alexandria

STATE: Virginia

COUNTRY: United States

ZIP: 22313-1404

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/091,028A

FILING DATE: 14-JUL-1993  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: JP 4-212305  
 FILING DATE: 17-JUL-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: JP 5-067339  
 FILING DATE: 04-MAR-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: REA, TERESA STANER  
 REGISTRATION NUMBER: 30,427  
 REFERENCE/DOCKET NUMBER: 001560-204  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 836-6620  
 TELEFAX: (703) 836-6620  
 INFORMATION FOR SEQ ID NO: 34:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 380 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-091-028A-34

Query Match 100.0%; Score 1959; DB 4; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-179;  
 Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASLAANAEFCNLFREMDNNGNNGVFFSSLSLFAALALVRLGAODDSLQIDKLLHV 60  
 DB 1 MASLAANAEFCNLFREMDNNGNNGVFFSSLSLFAALALVRLGAODDSLQIDKLLHV 60  
 QY 61 NTASGYGNSNSGSGLOSLKRVFSDINASHKDYDLSYNGLFAEKVYGFHKDYIECAEK 120  
 DB 61 NTASGYGNSNSGSGLOSLKRVFSDINASHKDYDLSYNGLFAEKVYGFHKDYIECAEK 120  
 QY 121 LYDAKVERVDFTNHLDETNRNINKWVENETHGKIKKNIYGGGJSSSAVMVLNAVYFKGK 180  
 DB 121 LYDAKVERVDFTNHLDETNRNINKWVENETHGKIKKNIYGGGJSSSAVMVLNAVYFKGK 180  
 QY 181 WQSAFTKSETINCHFKSPKCSGKAVAMHOERKFNLSVIEDPSMKLLELRNGIMNYVL 240  
 DB 181 WQSAFTKSETINCHFKSPKCSGKAVAMHOERKFNLSVIEDPSMKLLELRNGIMNYVL 240  
 QY 241 LPENDLSEIENKLTFFONLMEWTPRRMTSKYVEVEFPQKIEKNYEMKQYLRALGLKIDF 300  
 DB 241 LPENDLSEIENKLTFFONLMEWTPRRMTSKYVEVEFPQKIEKNYEMKQYLRALGLKIDF 300  
 QY 301 DESKADLSGIASGGRLYISRMHKSXYIEVEEGTEATAAGSNIVEKQLPOSTLFRADHP 360  
 DB 301 DESKADLSGIASGGRLYISRMHKSXYIEVEEGTEATAAGSNIVEKQLPOSTLFRADHP 360  
 QY 361 FLFVIRKDDIILFSGKVSCP 380  
 DB 361 FLFVIRKDDIILFSGKVSCP 380

RESULT 2  
 US-09-140-719-34  
 Sequence 34, Application us/09140719  
 GENERAL INFORMATION:  
 APPLICANT: TSUJIMOTO, Masafumi  
 APPLICANT: IMASA, Fuyuki  
 APPLICANT: TSURODOKA, Nobuo  
 APPLICANT: NAKAZATO, Hiroshi  
 APPLICANT: MURA, Kenju  
 APPLICANT: ISHIDA, Nobuhito  
 APPLICANT: KURIHARA, Tatsuya  
 APPLICANT: YAMACHI, Kozo  
 APPLICANT: YAMAGUCHI, Nozomi  
 TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
 NUMBER OF SEQUENCES: 34  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Burns, Doane, Swecker & Mathis

STREET: P.O. Box 1404  
 CITY: Alexandria  
 STATE: Virginia  
 COUNTRY: United States  
 ZIP: 22313-1404  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/140,719  
 FILING DATE: 08-AUG-1998  
 CLASSIFICATION:  
 APPLICATION DATA:  
 APPLICATION NUMBER: US 08/474,661  
 FILING DATE: 07-JUN-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/091,028  
 FILING DATE: 14-JUL-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: JP 4-212305  
 FILING DATE: 17-JUL-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: JP 6-067339  
 FILING DATE: 04-MAR-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: McGowan, Malcolm K.  
 REGISTRATION NUMBER: 39,300  
 REFERENCE/DOCKET NUMBER: 001560-247  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 836-6620  
 TELEFAX: (703) 836-2021  
 INFORMATION FOR SEQ ID NO: 34:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 380 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-09-140-719-34

Query Match 100.0%; Score 1959; DB 15; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-179;  
 Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASLAANAEFCNLFREMDNNGNNGVFFSSLSLFAALALVRLGAODDSLQIDKLLHV 60  
 DB 1 MASLAANAEFCNLFREMDNNGNNGVFFSSLSLFAALALVRLGAODDSLQIDKLLHV 60  
 QY 61 NTASGYGNSNSGSGLOSLKRVFSDINASHKDYDLSYNGLFAEKVYGFHKDYIECAEK 120  
 DB 61 NTASGYGNSNSGSGLOSLKRVFSDINASHKDYDLSYNGLFAEKVYGFHKDYIECAEK 120  
 QY 121 LYDAKVERVDFTNHLDETNRNINKWVENETHGKIKKNIYGGGJSSSAVMVLNAVYFKGK 180  
 DB 121 LYDAKVERVDFTNHLDETNRNINKWVENETHGKIKKNIYGGGJSSSAVMVLNAVYFKGK 180  
 QY 181 WQSAFTKSETINCHFKSPKCSGKAVAMHOERKFNLSVIEDPSMKLLELRNGIMNYVL 240  
 DB 181 WQSAFTKSETINCHFKSPKCSGKAVAMHOERKFNLSVIEDPSMKLLELRNGIMNYVL 240  
 QY 241 LPENDLSEIENKLTFFONLMEWTPRRMTSKYVEVEFPQKIEKNYEMKQYLRALGLKIDF 300  
 DB 241 LPENDLSEIENKLTFFONLMEWTPRRMTSKYVEVEFPQKIEKNYEMKQYLRALGLKIDF 300  
 QY 301 DESKADLSGIASGGRLYISRMHKSXYIEVEEGTEATAAGSNIVEKQLPOSTLFRADHP 360  
 DB 301 DESKADLSGIASGGRLYISRMHKSXYIEVEEGTEATAAGSNIVEKQLPOSTLFRADHP 360  
 QY 361 FLFVIRKDDIILFSGKVSCP 380  
 DB 361 FLFVIRKDDIILFSGKVSCP 380

RESULT 3  
US-09-508-997A-2  
Sequence 2, Application US/09508997A  
GENERAL INFORMATION:  
APPLICANT: MIYATA, Toshio  
TITLE OF INVENTION: Megsin Protein  
FILE REFERENCE: SHIM-004  
CURRENT APPLICATION NUMBER: US/09/508, 997A  
CURRENT FILING DATE: 2000-06-27  
PRIOR APPLICATION NUMBER: JP 09-275302  
PRIOR FILING DATE: 1997-09-22  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 2  
LENGTH: 380  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-508-997A-2

Query Match 100.0%; Score 1959; DB 19; Length 380;  
Best Local Similarity 100.0%; Pred. No. 1,1e-179;  
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASLAANAEEFCFNLFREMDNNGNNGVFFSSLSLFAALALVRLGADDSLSQIDKLLHV 60  
DB 1 MASLAANAEEFCFNLFREMDNNGNNGVFFSSLSLFAALALVRLGADDSLSQIDKLLHV 60  
QY 61 NTASGYGSSNSGSGLOSOLKRVFSDINASHKDYDLSIVNGLFAEKYGGFHKDYIECAEK 120  
DB 61 NTASGYGSSNSGSGLOSOLKRVFSDINASHKDYDLSIVNGLFAEKYGGFHKDYIECAEK 120  
QY 121 LYDAKVERVDFPTNLEPTRRINKVENEHGTIKNYIGEGGSSAVMVLVNAVYFRKG 180  
DB 121 LYDAKVERVDFPTNLEPTRRINKVENEHGTIKNYIGEGGSSAVMVLVNAVYFRKG 180  
QY 181 WSAFTKSETINCHFKSPKCSGKAVAMHDERKFNLSVIEDPSKILELRNGGINNYVL 240  
DB 181 WSAFTKSETINCHFKSPKCSGKAVAMHDERKFNLSVIEDPSKILELRNGGINNYVL 240  
QY 241 LPENDLSEIEKTKLFQNLMEWTNPRMTSKYVEVFFPOFKIEKYEKKOYLALGLDIF 300  
DB 241 LPENDLSEIEKTKLFQNLMEWTNPRMTSKYVEVFFPOFKIEKYEKKOYLALGLDIF 300  
QY 301 DESADLSGASGRLYISRMHKSYLEVTEEGTEATAAGSNIVEKQLPOSTLFRADHP 360  
DB 301 DESADLSGASGRLYISRMHKSYLEVTEEGTEATAAGSNIVEKQLPOSTLFRADHP 360  
QY 361 FLFVIRKDDIILFSGKVSCLP 380  
DB 361 FLFVIRKDDIILFSGKVSCLP 380

RESULT 4  
US-09-791-537-281  
Sequence 281, Application US/09791537  
GENERAL INFORMATION:  
APPLICANT: Biomimix, Inc.  
APPLICANT: Debe, Derek  
APPLICANT: Danzer, Joseph  
TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS  
FILE REFERENCE: 261/210  
CURRENT APPLICATION NUMBER: US/09/791, 537  
CURRENT FILING DATE: 2001-02-22  
NUMBER OF SEQ ID NOS: 153055  
SOFTWARE: Patent In Ver. 3.0  
SEQ ID NO 281  
LENGTH: 380  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-791-537-281

Query Match 100.0%; Score 1959; DB 21; Length 380;  
Best Local Similarity 100.0%; Pred. No. 1,1e-179;  
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASLAANAEEFCFNLFREMDNNGNNGVFFSSLSLFAALALVRLGADDSLSQIDKLLHV 60  
DB 1 MASLAANAEEFCFNLFREMDNNGNNGVFFSSLSLFAALALVRLGADDSLSQIDKLLHV 60  
QY 61 NTASGYGSSNSGSGLOSOLKRVFSDINASHKDYDLSIVNGLFAEKYGGFHKDYIECAEK 120  
DB 61 NTASGYGSSNSGSGLOSOLKRVFSDINASHKDYDLSIVNGLFAEKYGGFHKDYIECAEK 120  
QY 121 LYDAKVERVDFPTNLEPTRRINKVENEHGTIKNYIGEGGSSAVMVLVNAVYFRKG 180  
DB 121 LYDAKVERVDFPTNLEPTRRINKVENEHGTIKNYIGEGGSSAVMVLVNAVYFRKG 180  
QY 181 WSAFTKSETINCHFKSPKCSGKAVAMHDERKFNLSVIEDPSKILELRNGGINNYVL 240  
DB 181 WSAFTKSETINCHFKSPKCSGKAVAMHDERKFNLSVIEDPSKILELRNGGINNYVL 240  
QY 241 LPENDLSEIEKTKLFQNLMEWTNPRMTSKYVEVFFPOFKIEKYEKKOYLALGLDIF 300  
DB 241 LPENDLSEIEKTKLFQNLMEWTNPRMTSKYVEVFFPOFKIEKYEKKOYLALGLDIF 300  
QY 301 DESADLSGASGRLYISRMHKSYLEVTEEGTEATAAGSNIVEKQLPOSTLFRADHP 360  
DB 301 DESADLSGASGRLYISRMHKSYLEVTEEGTEATAAGSNIVEKQLPOSTLFRADHP 360  
QY 361 FLFVIRKDDIILFSGKVSCLP 380  
DB 361 FLFVIRKDDIILFSGKVSCLP 380

RESULT 5  
US-09-936-883A-2  
Sequence 2, Application US/09936883A  
GENERAL INFORMATION:  
APPLICANT: MIYATA, Toshio  
TITLE OF INVENTION: A method for Detecting Megsin Protein and Use  
FILE REFERENCE: F2-101DPICT  
CURRENT APPLICATION NUMBER: US/09/936, 883A  
CURRENT FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: JP 1999-75305  
PRIOR FILING DATE: 1999-03-19  
PRIOR APPLICATION NUMBER: JP 1999-306623  
PRIOR FILING DATE: 1999-10-28  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 2  
LENGTH: 380  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-936-883A-2

Query Match 100.0%; Score 1959; DB 23; Length 380;  
Best Local Similarity 100.0%; Pred. No. 1,1e-179;  
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASLAANAEEFCFNLFREMDNNGNNGVFFSSLSLFAALALVRLGADDSLSQIDKLLHV 60  
DB 1 MASLAANAEEFCFNLFREMDNNGNNGVFFSSLSLFAALALVRLGADDSLSQIDKLLHV 60  
QY 61 NTASGYGSSNSGSGLOSOLKRVFSDINASHKDYDLSIVNGLFAEKYGGFHKDYIECAEK 120  
DB 61 NTASGYGSSNSGSGLOSOLKRVFSDINASHKDYDLSIVNGLFAEKYGGFHKDYIECAEK 120  
QY 121 LYDAKVERVDFPTNLEPTRRINKVENEHGTIKNYIGEGGSSAVMVLVNAVYFRKG 180  
DB 121 LYDAKVERVDFPTNLEPTRRINKVENEHGTIKNYIGEGGSSAVMVLVNAVYFRKG 180  
QY 181 WSAFTKSETINCHFKSPKCSGKAVAMHDERKFNLSVIEDPSKILELRNGGINNYVL 240  
DB 181 WSAFTKSETINCHFKSPKCSGKAVAMHDERKFNLSVIEDPSKILELRNGGINNYVL 240

Db 181 WSAFTKSETINCHFKSPKSGKAVAMMOERKFNLSVIEDPSMKLLELRNGGIMMYVL 240  
 QY 241 LPENDLSEIENKLTFFONLMEWTNPRMTSKYVEFFPPQKIEKNYEMKOYLALGKIDIF 300  
 Db 241 LPENDLSEIENKLTFFONLMEWTNPRMTSKYVEFFPPQKIEKNYEMKOYLALGKIDIF 300  
 QY 301 DESKADLSGIASGGRLYISRMHKSXYIEVTEEGTEATAATGSNIYVKOLPOSTLFRADHP 360  
 Db 301 DESKADLSGIASGGRLYISRMHKSXYIEVTEEGTEATAATGSNIYVKOLPOSTLFRADHP 360  
 QY 361 FLFVIRKDDIILFSGKVSCEP 380  
 Db 361 FLFVIRKDDIILFSGKVSCEP 380

## RESULT 6

US-10-091-442-34

Sequence 34, Application US/10091442

GENERAL INFORMATION:

APPLICANT: TSUJIMOTO, Masafumi

IMASA, Fuyuki

TSUROUOKA, Nobuo

NAKAZATO, Hiroshi

MURA, Kenju

ISHIDA, Nobuhiko

KURIHARA, Tatsuya

YAMAICHI, Kozo

YAMAGUCHI, Nozomi

TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR

NUMBER OF SEQUENCES: 34

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Burns, Doane, Swecker &amp; Mathis

STREET: P.O. Box 1404

CITY: Alexandria

STATE: Virginia

COUNTRY: United States

ZIP: 22313-1404

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/091,442

FILING DATE: 07-Mar-2002

CLASSIFICATION: &lt;Unknown&gt;

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/140,719

FILING DATE: 08-Aug-1998

APPLICATION NUMBER: US 08/474,661

FILING DATE: 07-JUN-1995

APPLICATION NUMBER: US 08/091,028

FILING DATE: 14-JUL-1993

APPLICATION NUMBER: JP 4-212305

FILING DATE: 17-JUL-1992

APPLICATION NUMBER: JP 6-067339

FILING DATE: 04-MAR-1993

ATTORNEY/AGENT INFORMATION:

NAME: McGowan, Malcolm K.

REGISTRATION NUMBER: 39,300

REFERENCE/DOCKET NUMBER: 001560-247

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 836-6620

TELEFAX: (703) 836-2021

INFORMATION FOR SEQ ID NO: 34:

SEQUENCE CHARACTERISTICS:

LENGTH: 380 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 34:

US-10-091-442-34

Query Match 100.0%; Score 1959; DB 24; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 1,1e-179;  
 Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASLAANAEEFCNLFREDDDNGNGNVFFSSLSLFAALALVRLGQODSLSDIDKLHV 60  
 Db 1 MASLAANAEEFCNLFREDDDNGNGNVFFSSLSLFAALALVRLGQODSLSDIDKLHV 60  
 QY 61 NTASGYGNSNSGSLQSLKRVFSDINASHKDYDLSYNGLFAEKVYGFHKDYIECAEK 120  
 Db 61 NTASGYGNSNSGSLQSLKRVFSDINASHKDYDLSYNGLFAEKVYGFHKDYIECAEK 120  
 QY 121 LYDAKVERVDFTNHLIEDTRRNINKWVENETHGKIKNVIGEGGSSSAVVLVNAVYFKGK 180  
 Db 121 LYDAKVERVDFTNHLIEDTRRNINKWVENETHGKIKNVIGEGGSSSAVVLVNAVYFKGK 180  
 QY 181 WSAFTKSETINCHFKSPKSGKAVAMMOERKFNLSVIEDPSMKLLELRNGGIMMYVL 240  
 Db 181 WSAFTKSETINCHFKSPKSGKAVAMMOERKFNLSVIEDPSMKLLELRNGGIMMYVL 240  
 QY 241 LPENDLSEIENKLTFFONLMEWTNPRMTSKYVEFFPPQKIEKNYEMKOYLALGKIDIF 300  
 Db 241 LPENDLSEIENKLTFFONLMEWTNPRMTSKYVEFFPPQKIEKNYEMKOYLALGKIDIF 300  
 QY 301 DESKADLSGIASGGRLYISRMHKSXYIEVTEEGTEATAATGSNIYVKOLPOSTLFRADHP 360  
 Db 301 DESKADLSGIASGGRLYISRMHKSXYIEVTEEGTEATAATGSNIYVKOLPOSTLFRADHP 360  
 QY 361 FLFVIRKDDIILFSGKVSCEP 380  
 Db 361 FLFVIRKDDIILFSGKVSCEP 380

## RESULT 7

US-10-110-030-2

Sequence 2, Application US/10110030

GENERAL INFORMATION:

APPLICANT: Miyata, Toshio

TITLE OF INVENTION: ANIMAL MODEL FOR MESENCHYMAL PROLIFERATIVE

TITLE OF INVENTION: NEPHRITIS

FILE REFERENCE: SHIM015

CURRENT APPLICATION NUMBER: US/10/110,030

CURRENT FILING DATE: 2002-04-03

PRIOR APPLICATION NUMBER: 11/285736

PRIOR FILING DATE: 1999-10-06

NUMBER OF SEQ ID NOS: 15

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO: 2

LENGTH: 380

TYPE: PRT

ORGANISM: Homo sapiens

US-10-110-030-2

Query Match 100.0%; Score 1959; DB 25; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 1,1e-179;  
 Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASLAANAEEFCNLFREDDDNGNGNVFFSSLSLFAALALVRLGQODSLSDIDKLHV 60  
 Db 1 MASLAANAEEFCNLFREDDDNGNGNVFFSSLSLFAALALVRLGQODSLSDIDKLHV 60  
 QY 61 NTASGYGNSNSGSLQSLKRVFSDINASHKDYDLSYNGLFAEKVYGFHKDYIECAEK 120  
 Db 61 NTASGYGNSNSGSLQSLKRVFSDINASHKDYDLSYNGLFAEKVYGFHKDYIECAEK 120  
 QY 121 LYDAKVERVDFTNHLIEDTRRNINKWVENETHGKIKNVIGEGGSSSAVVLVNAVYFKGK 180  
 Db 121 LYDAKVERVDFTNHLIEDTRRNINKWVENETHGKIKNVIGEGGSSSAVVLVNAVYFKGK 180  
 QY 181 WSAFTKSETINCHFKSPKSGKAVAMMOERKFNLSVIEDPSMKLLELRNGGIMMYVL 240  
 Db 181 WSAFTKSETINCHFKSPKSGKAVAMMOERKFNLSVIEDPSMKLLELRNGGIMMYVL 240



Oy 241 LPENDLSEIENKLTFFONLMEWTNPRMTSKYVEVFFPOFKTEKVENKQYLALGLKIDIF 300  
Db 241 LPENDLSEIENKLTFFONLMEWTNPRMTSKYVEVFFPOFKTEKVENKQYLALGLKIDIF 300  
Oy 301 DESRADLSGIASGRLYISRMHKSYLEVTEEGTEATATGSNIYEVKOLPOSTLFRADHP 360  
Db 301 DESRADLSGIASGRLYISRMHKSYLEVTEEGTEATATGSNIYEVKOLPOSTLFRADHP 360  
Oy 361 FLFVIRKDDITLFGSKVSCP 380  
Db 361 FLFVIRKDDITLFGSKVSCP 380

## RESULT 8

US-09-508-997A-4  
Sequence 4, Application US/09508997A  
GENERAL INFORMATION:  
APPLICANT: MIYATA, Toshio  
TITLE OF INVENTION: Megsin Protein  
FILE REFERENCE: SHIM-004  
CURRENT APPLICATION NUMBER: US/09/508,997A  
CURRENT FILING DATE: 2000-06-27  
PRIOR APPLICATION NUMBER: JP 09-275302  
PRIOR FILING DATE: 1997-09-22  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 4  
LENGTH: 380  
TYPE: PRT  
ORGANISM: Rattus rattus  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: 51, 94  
OTHER INFORMATION: Xaa is unknown.  
US-09-508-997A-4

Query Match 75.4%; Score 1477; DB 19; Length 380;  
Best Local Similarity 73.4%; Pred. No. 3,8e-133;  
Matches 279; Conservative 56; Mismatches 45; Indels 0; Gaps 0;

Oy 1 MASLAANAEEFCNLFREMDNNGNNGVFFSSLSLFAALALVRLGADDSLSQIDKLLHV 60  
Db 1 MASLAANAEEFCNLFREMDNNGNNGVFFSSLSLFAALALVRLGADDSLSQIDKLLHV 60  
Oy 61 NTASGYGNSNSQSGLOLQKRVFSDINASHKDDVLSVNGLFAEKYVGFHKKDYIECAEK 120  
Db 61 ISPSRQGNSSNSQSGLOLQKRVFSDINASHKDDVLSVNGLFAEKYVGFHKKDYIECAEK 120  
Oy 121 LYDAKVERVDFTNHLIEDTRNINRWENETHGKIKNVIGEGISSAAVMVLVNAVYFEGK 180  
Db 121 LYNAKVERVDFTNHLIEDTRNINRWENETHGKIKNVIGEGISSAAVMVLVNAVYFEGK 180  
Oy 181 WQSAFTSETINCHFKSPKSCGKAVAMMHORRKNLSYIEDPSMKLIELRYNGGINMYVL 240  
Db 181 WQSAFTSETINCHFKSPKSCGKAVAMMHORRKNLSYIEDPSMKLIELRYNGGINMYVL 240  
Oy 241 LPENDLSEIENKLTFFONLMEWTNPRMTSKYVEVFFPOFKTEKVENKQYLALGLKIDIF 300  
Db 241 LPENDLSEIENKLTFFONLMEWTNPRMTSKYVEVFFPOFKTEKVENKQYLALGLKIDIF 300  
Oy 301 DESRADLSGIASGRLYISRMHKSYLEVTEEGTEATATGSNIYEVKOLPOSTLFRADHP 360  
Db 301 DESRADLSGIASGRLYISRMHKSYLEVTEEGTEATATGSNIYEVKOLPOSTLFRADHP 360  
Oy 361 FLFVIRKDDITLFGSKVSCP 380  
Db 361 FLFVIRKDDITLFGSKVSCP 380

RESULT 9  
US-09-936-883A-19  
Sequence 19, Application US/09936883A  
GENERAL INFORMATION:

APPLICANT: MIYATA, Toshio  
TITLE OF INVENTION: A Method for Detecting Megsin Protein and Use  
FILE REFERENCE: F2-101DPICT  
CURRENT APPLICATION NUMBER: US/09/936,883A  
CURRENT FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: JP 1999-75305  
PRIOR FILING DATE: 1999-03-19  
PRIOR APPLICATION NUMBER: JP 1999-10-28  
PRIOR FILING DATE: 1999-10-28  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 19  
LENGTH: 380  
TYPE: PRT  
ORGANISM: Rattus norvegicus  
FEATURE:  
NAME/KEY: misc-feature  
LOCATION: 51, 94  
OTHER INFORMATION: Xaa is unknown.  
US-09-936-883A-19

Query Match 75.4%; Score 1477; DB 23; Length 380;  
Best Local Similarity 73.4%; Pred. No. 3,8e-133;  
Matches 279; Conservative 56; Mismatches 45; Indels 0; Gaps 0;

Oy 1 MASLAANAEEFCNLFREMDNNGNNGVFFSSLSLFAALALVRLGADDSLSQIDKLLHV 60  
Db 1 MASLAANAEEFCNLFREMDNNGNNGVFFSSLSLFAALALVRLGADDSLSQIDKLLHV 60  
Oy 61 NTASGYGNSNSQSGLOLQKRVFSDINASHKDDVLSVNGLFAEKYVGFHKKDYIECAEK 120  
Db 61 ISPSRQGNSSNSQSGLOLQKRVFSDINASHKDDVLSVNGLFAEKYVGFHKKDYIECAEK 120  
Oy 121 LYDAKVERVDFTNHLIEDTRNINRWENETHGKIKNVIGEGISSAAVMVLVNAVYFEGK 180  
Db 121 LYNAKVERVDFTNHLIEDTRNINRWENETHGKIKNVIGEGISSAAVMVLVNAVYFEGK 180  
Oy 181 WQSAFTSETINCHFKSPKSCGKAVAMMHORRKNLSYIEDPSMKLIELRYNGGINMYVL 240  
Db 181 WQSAFTSETINCHFKSPKSCGKAVAMMHORRKNLSYIEDPSMKLIELRYNGGINMYVL 240  
Oy 241 LPENDLSEIENKLTFFONLMEWTNPRMTSKYVEVFFPOFKTEKVENKQYLALGLKIDIF 300  
Db 241 LPENDLSEIENKLTFFONLMEWTNPRMTSKYVEVFFPOFKTEKVENKQYLALGLKIDIF 300  
Oy 301 DESRADLSGIASGRLYISRMHKSYLEVTEEGTEATATGSNIYEVKOLPOSTLFRADHP 360  
Db 301 DESRADLSGIASGRLYISRMHKSYLEVTEEGTEATATGSNIYEVKOLPOSTLFRADHP 360  
Oy 361 FLFVIRKDDITLFGSKVSCP 380  
Db 361 FLFVIRKDDITLFGSKVSCP 380

RESULT 10  
US-10-110-030-4  
Sequence 4, Application US/10110030  
GENERAL INFORMATION:  
APPLICANT: Miyata, Toshio  
TITLE OF INVENTION: ANIMAL MODEL FOR MESENCHYMAL PROLIFERATIVE  
FILE REFERENCE: NEPHRITS  
CURRENT APPLICATION NUMBER: US/10/110,030  
CURRENT FILING DATE: 2002-04-03  
PRIOR APPLICATION NUMBER: 11/285736  
PRIOR FILING DATE: 1999-10-06  
NUMBER OF SEQ ID NOS: 15  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 4  
LENGTH: 380  
TYPE: PRT  
ORGANISM: Rattus norvegicus

FEATURE:  
 NAME/KEY: VARIANT  
 LOCATION: 51, 94  
 OTHER INFORMATION: Xaa - Any Amino Acid  
 US-10-110-030-4

Query Match 75.4%; Score 1477; DB 25; Length 380;  
 Best Local Similarity 73.4%; Pred. No. 3.8e-133;  
 Matches 279; Conservative 56; Mismatches 45; Indels 0; Gaps 0;

QY 1 MASLAANAEFCNLTREMNOGNGVNFSSLSLFAALALVRLGAODDSLQIDKLHVNTASGNSNS 60  
 D 1 MASLAANAEFGDLREMDSSOGNGVNFSSLSLFTALSLIRLGARDCXQIDALH 60  
 QY 61 NTASGYGNSNSOGSLQSLKRVFSDINASHKDYDLSIVNGLFAEKYGFHKDYIECAEK 120  
 D 61 ISPSROGNSNSOGSLQSLQYOLKRVLADINSHKDKXLSIANGVFAEKYFEDHKRYMECAEN 120  
 QY 121 LIDAKYERVDFTNHLDETRNINKNVVENETHGKIKVNGIGGSISSAVMLVNAVYFKG 180  
 D 121 LYNAKERYVDFTNIDIOETREKIKVNIENETHGKIKVLDGSSSSAVMLVNAVYFKG 180  
 QY 181 WSAFRTSEITNCHFPKSGKAVAMHOERKFNLSYEDPSMKITELRYNGINMYVL 240  
 D 181 WSAFRTKSDTSLCHFPSPGKAVAMHOERFNLSIOEPMOILEQYHGGISMYIM 240  
 QY 241 LPENDLSELENKLTFFONLMEWTNPRMTSKYVEVFPOFKIEKNYEMKOYLALGLKIDIF 300  
 D 241 LPEDDLSEIESKLSFQNLMDWTNSRKMKSOYAVFLPQFLEIENDYERSHLSKVGLEDIF 300  
 QY 301 DESKADLSIASGRLYISRMHKSYLEVEEGTEATATGSNIVEKOLPOSTLFRADHP 360  
 D 301 VESRADLSIASGRLYISRMHKSILEVEEGTEATATGSNIVEKLPSTVFRADRP 360  
 QY 361 FLFVIRKDDTLFSGVSCP 380  
 D 361 FLFVIRKNGITLFTGVSCP 380

## RESULT 11.

US-09-508-997A-6  
 Sequence 6, Application US/09508997A  
 GENERAL INFORMATION:  
 APPLICANT: MIYATA, Toshio  
 TITLE OF INVENTION: Megin Protein  
 FILE REFERENCE: SHIM-004  
 CURRENT APPLICATION NUMBER: US/09/508, 997A  
 PRIOR FILING DATE: 2000-06-27  
 PRIOR APPLICATION NUMBER: JP 09-275302  
 NUMBER OF SEQ ID NOS: 44  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 6  
 LENGTH: 368  
 TYPE: PRT  
 ORGANISM: Mus musculus  
 US-09-508-997A-6

Query Match 74.4%; Score 1458; DB 19; Length 368;  
 Best Local Similarity 73.4%; Pred. No. 2.5e-131;  
 Matches 270; Conservative 53; Mismatches 45; Indels 0; Gaps 0;

QY 13 FNLFREMDNOGNGVNFSSLSLFAALALVRLGAODDSLQIDKLHVNTASGNSNS 72  
 D 1 FDLFREMDSSOGNGVNFSSLSLFTALTILIRLGARDCARQIDKALHFNIPSGNSNN 60  
 QY 73 QSGLSQSLKRVFSDINASHKDYDLSIVNGLFAEKYGFHKDYIECAEKLYDAKVERVDFT 132  
 D 61 QGQLQYOLKRVLADINSHKDYELSTATGVFAEKYDFHKNYIECAENLYNKAVERVDFT 120  
 QY 133 NHLEPTRNINKNVVENETHGKIKVNGIGGSISSAVMLVNAVYFGKQSAFTKSETIN 192  
 D 121 NDVQDTRFKINKNVNIENETHGKIKVLDGSSSSAVMLVNAVYFGKQSAFTKTDTL 180

QY 193 CHEFSPKSGKAVAMHOERKFNLSYEDPSMKITELRYNGINMYVLPENDLSELENK 252  
 D 181 CRFRSPCTCKKYVNMHMOERFNLSIQPRMAYLEQYHGGISWIMLPEDDLCEIEBK 240  
 QY 253 LTFONLMEWTNPRMTSKYVEVFPOFKIEKNYEMKOYLALGLKIDFESKADLSGAS 312  
 D 241 LSFQNLMDWTNRRKMKSOYAVFLPQFLEIENDYERSHLSKVGLEDIFESSADLSGAS 300  
 QY 313 GGRLYISRMHKSYLEVEEGTEATATGSNIVEKOLPOSTLFRADHPFLYIRKDDITL 372  
 D 301 GGRLYYSKLMHKSILEVEEGTEATATGSNIVEKOLPESTVFRADRPFLYIKKDDITL 360  
 QY 373 FSGKAVSCP 380  
 D 361 FTGKAVSCP 368

## RESULT 12

US-09-936-883A-21  
 Sequence 21, Application US/09936883A  
 GENERAL INFORMATION:

APPLICANT: MIYATA, Toshio  
 TITLE OF INVENTION: A Method for Detecting Megin Protein and Use  
 FILE REFERENCE: F2-101DIPCT  
 CURRENT APPLICATION NUMBER: US/09/936, 883A  
 PRIOR FILING DATE: 2001-12-21  
 PRIOR APPLICATION NUMBER: JP 1999-75305  
 PRIOR FILING DATE: 1999-03-19  
 PRIOR APPLICATION NUMBER: JP 1999-306623  
 NUMBER OF SEQ ID NOS: 21  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 21  
 LENGTH: 368  
 TYPE: PRT  
 ORGANISM: Mus musculus  
 US-09-936-883A-21

Query Match 74.4%; Score 1458; DB 23; Length 368;  
 Best Local Similarity 73.4%; Pred. No. 2.5e-131;  
 Matches 270; Conservative 53; Mismatches 45; Indels 0; Gaps 0;

QY 13 FNLFREMDNOGNGVNFSSLSLFAALALVRLGAODDSLQIDKLHVNTASGNSNS 72  
 D 1 FDLFREMDSSOGNGVNFSSLSLFTALTILIRLGARDCARQIDKALHFNIPSGNSNN 60  
 QY 73 QSGLSQSLKRVFSDINASHKDYDLSIVNGLFAEKYGFHKDYIECAEKLYDAKVERVDFT 132  
 D 61 QGQLQYOLKRVLADINSHKDYELSTATGVFAEKYDFHKNYIECAENLYNKAVERVDFT 120  
 QY 133 NHLEPTRNINKNVVENETHGKIKVNGIGGSISSAVMLVNAVYFGKQSAFTKSETIN 192  
 D 121 NDVQDTRFKINKNVNIENETHGKIKVLDGSSSSAVMLVNAVYFGKQSAFTKTDTL 180  
 QY 193 CHEFSPKSGKAVAMHOERKFNLSYEDPSMKITELRYNGINMYVLPENDLSELENK 252  
 D 181 CRFRSPCTCKKYVNMHMOERFNLSIQPRMAYLEQYHGGISWIMLPEDDLCEIEBK 240  
 QY 253 LTFONLMEWTNPRMTSKYVEVFPOFKIEKNYEMKOYLALGLKIDFESKADLSGAS 312  
 D 241 LSFQNLMDWTNRRKMKSOYAVFLPQFLEIENDYERSHLSKVGLEDIFESSADLSGAS 300  
 QY 313 GGRLYISRMHKSYLEVEEGTEATATGSNIVEKOLPOSTLFRADHPFLYIRKDDITL 372  
 D 301 GGRLYYSKLMHKSILEVEEGTEATATGSNIVEKOLPESTVFRADRPFLYIKKDDITL 360  
 QY 373 FSGKAVSCP 380  
 D 361 FTGKAVSCP 368

```
RESULT 13
US-10-110-030-6
; Sequence 6, Application US/10110030
; GENERAL INFORMATION:
; APPLICANT: Miyata, Toshio
; TITLE OF INVENTION: ANIMAL MODEL FOR MESENCHYMAL PROLIFERATIVE
; FILE REFERENCE: SHIM015
; CURRENT APPLICATION NUMBER: US/10/110,030
; CURRENT FILING DATE: 2002-04-03
; PRIOR APPLICATION NUMBER: 11/285736
; PRIOR FILING DATE: 1999-10-06
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 368
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-110-030-6

Query Match          74.4% Score 1458: DB 25: Length 368:
Best Local Similarity 73.4%: Pred. No. 2.5e-131;
Matches 270; Conservative 53; Mismatches 45; Indels 0; Gaps 0;

QY 13 FNLREMDNDQNGNGNFFSSLSLFAALALVRLGAQDLSQIDKLHVTNATSGYNGSSNS 72
DB 1 FDLFREDDSSQNGNGNFFSSLSLFTALTRLRAGRCARQIDKALHFNIPRQNGSSNN 60
QY 73 QSGLSQSLKRVFSDINASHKDYLSTVNGLFAEKVGFPHKDYIECAEKLVDKVERVDF 132
DB 61 QPGLQYOLKRVLDADINSHKDYELSTVNGLFAEKVGFPHKDYIECAEKLVDKVERVDF 120
QY 133 NHELETRRNKNWENTHGRKIKVIGEGGSSSAVAVLVNAVYFKGKMSAFKSTPTN 192
DB 121 NDVODTRFKINKMTEHTHGRKIKVLDSSSSAVAVLVNAVYFKGKMSAFKSTPTLS 180
QY 193 CHERPSKCGKAAVAMHOKERKFNLSYEDPSMKLIELRVNGINMYVLLPENDINGSELEN 252
DB 181 CRFRSPTRCPGKVVNMHGERFNSTIQPRMQLDQYNGISMTIMLEDGCELESK 240
QY 253 LTFONLMEHTNPRRMTSKYVEFFPPQKTEKNYEMKQYLRALGLKIDFDESKADLSGIAS 312
DB 241 LSEFONLMDWTNRRKMKSOYVNVFLPQPKIEKNYEMTHLKLKIDFDESSADLSGIAS 300
QY 313 GGRLYISRMHKSYTEVTEGETATATGSIYVEKOPSTIRPADHPFLVIRKDDIIL 372
DB 301 GGRLYISKLHMKSTIEVEGETATATENNIVEKOLPESTIVERADHPFLVIRKDDIIL 360
QY 373 FSGKVSCEP 380
DB 361 FTGKVSCEP 368

RESULT 14
US-09-283-946-2
; Sequence 2, Application US/09283946
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel E.
; APPLICANT: Hubert, Rene S.
; APPLICANT: Leong, Kahen
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Saffran, Douglas C.
; TITLE OF INVENTION: PROSTAPIN GENE AND PROTEIN AND USES THEREOF
; FILE REFERENCE: 1703-005.US1
; CURRENT APPLICATION NUMBER: US/09/283,946
; CURRENT FILING DATE: 1999-04-01
; EARLIER APPLICATION NUMBER: 60/080,167
; EARLIER FILING DATE: 1998-03-31
; EARLIER APPLICATION NUMBER: 60/085,720
; EARLIER FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 2
```

```
LENGTH: 379
; TYPE: PRT
; ORGANISM: HUMAN PROSTAPIN CONSENSUS OF AMINO ACID SEQUENCE
US-09-283-946-2

Query Match          40.7% Score 796.5: DB 16: Length 379:
Best Local Similarity 42.6%: Pred. No. 2e-67;
Matches 168; Conservative 77; Mismatches 120; Indels 29; Gaps 8;

QY 1 MASIAAANAEFCENLFREMDNDQNGNGNFFSSLSLFAALALVRLGAQDLSQIDKLHAY 60
DB 1 MGSISTANVEFCIDVFEKELNSNNGDNIFFSSLSLVALSMVLGAGGETAEQLEKVLAF 60
QY 61 -----NTASGYNGSSN--SOSG-LQSOLKRVFSDINASHKDYDLSYNGLFAEKVGFPHK 113
DB 61 SHYVDSLKPFKDSKPSQAGRIHSEGFVSQINODPSNCTSIANRLIGTKTMAFHQ 120
QY 114 YIECAEKLVDKVERVDFNHELETRRNKNWENTHGRKIKVIGGGISSSAVAVLVN 173
DB 121 YLSCSEKWKQARLQTVDFEOSTEETRKIMANWENKTNKGVANLFGKSTIDPSSVAVLVN 180
QY 174 AVYFKGKWSAFKSTETINCHFKSPKSGK--AVAMHOKERKFNLSYIEDPSMKLIELRY 231
DB 181 TIYFKGGRONKF-----QGRNVYEMVYQIGTEFLARVKEPQOVLELPY 225
QY 232 -NGGINVYVLLPEN--DLSELENKLFONLMEHTNPRRMTSKYVEFFPPQKTEKNYEMK 288
DB 226 VNNKLSMIIILPVGINALVKQIEKOLNSGTHEHTSSNMAREVEVLLPREFLEIKYELN 285
QY 289 QYLRALGLKIDFDESKADLSGIASGRLYISRMHKSYTEVTEGETATATGSIYVEKQ 348
DB 286 SLKPLGVTDLDEQNVKRLDLSGMSPTKGLYSKAIHKSYLDVSEGTAAATGDSIAVKS 345
QY 349 LPQSTIRPADHPFLVIR--KDDIILFSGVSCP 380
DB 346 LPMRAQPKANHPFLFIRHTHTNTIILFCGLASP 379

RESULT 15
US-09-852-911-946
; Sequence 946, Application US/09852911
; GENERAL INFORMATION:
; APPLICANT: Xu, Jianshun
; APPLICANT: Dillon, David C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stoik, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasser A.W.
; APPLICANT: Hepler, William
; APPLICANT: Hurai, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C25
; CURRENT APPLICATION NUMBER: US/09/852,911
; CURRENT FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 947
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 946
; LENGTH: 379
; TYPE: PRT
; ORGANISM: Homo sapiens
```

US-09-852-911-946

Query Match 40.7%; Score 796.5; DB 22; Length 379;  
Best Local Similarity 42.6%; Pred. No. 2e-67;  
Matches 168; Conservative 77; Mismatches 120; Indels 29; Gaps 8;

```
QY      1 MASLAANAEEFCNLFRENDNOCNGNVFFSSLSLPAALALVRLGADDSLSQIDKLLHV 60
      1 MGSLSLANVEFCIDVFEKELNSNIGNIFFSSLSLLYALSMVLLGARGETAEOLEKVLHF 60
DB      61 -----NTASGYGNSSN-SOSG-LQSOLKRYESDINASHKDYDLSIVNGLFAEKVYGPBKD 113
      61 SHTVDSLKPFKSPKCSQAGRIHSEFGEVFEFSQINOPDSNCTLSIANRLGKTMAFHQO 120
QY      114 YTECAEKLYDAKVERVFTNHLIEDTRRNINKWVENETHGKIKNVIGEGISSAVMVLVN 173
      121 YLSCSEKWTQARLQTVDFEOSTEETRKMINAWENTKNGKANLFGKSTIDPSSVMVLVN 180
DB      174 AVYFKGKQSAFTKSETINCHFKSPKCSGK--AVAMHQRKFNLSVIEDPSMKILELRY 231
      181 TIYFKGQRONKF-----OGKNVTVEWMYQIGTEKLAFAVKEPQMQLLELPY 225
QY      232 -NGGIMVYLLPEN--DLSEIENKLFQNLMEWTNRRMTSKYVEVEFFQFKTEKNYEMK 288
      226 VNNKLSMIILLPEVGIANLKQIEKQLNSGTFHEWTSSNMEREVEVHLPRFKLEIKYEELN 285
QY      289 QYLRALGLKDIEDSKADLSGIASGRLYISRMHKSYLEVEEGTEATATGSNIYVKO 348
      286 SLTKPLGVYDLEFNQVADLSGMSPTKGLYLSKAIHKSYYLDVSEEGTEAATAATGDSIAVKS 345
DB      349 LPOSTLFRADHPLEVIR--KDDIILFSGKVSCP 380
      346 LPMRAQFKANHPLEFIRHHTWTILFCGKLASP 379
```

Search completed: July 11, 2003, 12:27:41  
Job time : 1197.09 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:55:27 ; Search time 357.089 Seconds  
(without alignments)  
280.699 Million cell updates/sec

Title: US-10-091-442-34  
Perfect score: 1959  
Sequence: 1 MASLAANAEEFCNLFREMD.....FLFVIRKDDILFSGKVSQP 380

Scoring table: BLOSUM62  
Gap: 10.0, Gapext 0.5

Searched: 1229476 seqs, 263775074 residues

Total number of hits satisfying chosen parameters: 1229476

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 100%  
Maximum Match 100%  
Listing first 45 summaries

Database:

Pending-Patents\_AA\_New:\*  
1: /cgn2\_6/ptodata/1/paa/PCT\_NEW\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/paa/PCT\_NEW\_COMB.pep4:\*  
3: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep4:\*  
5: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep4:\*  
7: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep:\*  
8: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep4:\*  
9: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep:\*  
10: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep4:\*  
11: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep:\*  
12: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep4:\*  
13: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep:\*  
14: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1959	100.0	380	10 US-09-936-883C-2	Sequence 2, Appl
2	1959	100.0	380	12 US-10-126-052A-263	Sequence 263, App
3	1959	100.0	380	14 US-60-443-566-3006	Sequence 3006, Ap
4	1959	100.0	380	14 US-60-455-444-4692	Sequence 4692, Ap
5	1959	100.0	380	14 US-60-465-241-4692	Sequence 4692, Ap
6	1477	75.4	380	10 US-09-936-883C-19	Sequence 19, Appl
7	1458	74.4	368	10 US-09-936-883C-21	Sequence 21, Appl
8	822	42.0	372	14 US-60-443-566-3717	Sequence 3717, Ap
9	822	42.0	372	14 US-60-455-444-5630	Sequence 5630, Ap
10	822	42.0	372	14 US-60-465-241-5630	Sequence 5630, Ap
11	822	42.0	392	12 US-10-198-070-62	Sequence 62, Appl
12	820	41.9	392	12 US-10-198-070-67	Sequence 67, Appl
13	819	41.8	392	12 US-10-198-070-70	Sequence 70, Appl
14	796.5	40.7	379	12 US-10-294-025-946	Sequence 946, Appl
15	769	39.3	394	12 US-10-258-951-59	Sequence 59, Appl
16	763.5	39.0	405	2 PCT-US02-38526-416	Sequence 416, App
17	763.5	39.0	405	14 US-60-443-566-4407	Sequence 4407, Ap
18	763.5	39.0	405	14 US-60-455-444-7462	Sequence 7462, Ap
19	763.5	39.0	405	14 US-60-465-241-7462	Sequence 7462, Ap

20	754.5	38.5	425	2	PCT-US02-072158-2	Sequence 2, Appl1
21	754.5	38.5	425	12	US-10-419-277-2	Sequence 2, Appl1
22	754.5	38.5	425	14	US-60-443-566-4406	Sequence 4406, Ap
23	754.5	38.5	425	14	US-60-455-444-7461	Sequence 7461, Ap
24	754.5	38.5	425	14	US-60-465-241-7461	Sequence 7461, Ap
25	753.5	38.5	425	2	PCT-US02-38526-408	Sequence 408, App
26	752	38.4	390	2	PCT-US02-072158-12	Sequence 12, Appl
27	752	38.4	390	2	PCT-US03-17409-175	Sequence 175, Appl
28	752	38.4	390	14	US-60-448-285-5	Sequence 5, Appl1
29	751	38.3	410	10	US-09-724-676A-57391	Sequence 57391, A
30	751	38.3	410	10	US-09-724-676A-57391	Sequence 57391, A
31	748	38.2	390	2	PCT-US02-38526-417	Sequence 417, Appl
32	748	38.2	390	2	PCT-US02-072158-13	Sequence 13, Appl
33	748	38.2	390	12	US-10-126-052A-465	Sequence 465, App
34	748	38.2	390	12	US-10-295-027-177	Sequence 177, App
35	748	38.2	390	12	US-10-295-027-794	Sequence 794, App
36	748	38.2	390	14	US-60-448-285-6	Sequence 6, Appl1
37	738.5	37.7	617	2	PCT-US02-072158-14	Sequence 14, Appl
38	738.5	37.7	617	12	US-10-294-025-947	Sequence 947, App
39	722.5	36.9	379	2	PCT-US03-12731-14	Sequence 14, Appl
40	722.5	36.9	379	2	PCT-US02-072158-9	Sequence 9, Appl1
41	722.5	36.9	379	14	US-60-448-285-3	Sequence 3, Appl1
42	722.5	36.9	379	14	US-60-453-135-8341	Sequence 8341, Ap
43	722.5	36.9	379	14	US-60-453-050-8341	Sequence 8341, Ap
44	722.5	36.9	379	14	US-60-455-444-4629	Sequence 4629, Ap
45	722.5	36.9	379	14	US-60-465-241-4629	Sequence 4629, Ap

#### ALIGNMENTS

RESULT 1  
US-09-936-883C-2  
Sequence 2, Application US/09936883C  
GENERAL INFORMATION:  
APPLICANT: MIYATA, Toshio  
TITLE OF INVENTION: A Method for Detecting Megsin Protein and Use  
TITLE OF INVENTION: Thereof  
FILE REFERENCE: P2-101DIPCT  
CURRENT APPLICATION NUMBER: US/09/936, 883C  
CURRENT FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: JP 1999-75305  
PRIOR FILING DATE: 1999-03-19  
PRIOR APPLICATION NUMBER: JP 1999-306623  
PRIOR FILING DATE: 1999-10-28  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 380  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-936-883C-2

Query Match	Best Local Similarity	100.0%;	Score 1959;	DB 10;	Length 380;
Matches 380;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	1	MASLAANAEEFCNLFREMDNOCNGVFFSSLSLFAALALVRLGADDSLSQIDKLHV	60		
DB	1	MASLAANAEEFCNLFREMDNOCNGVFFSSLSLFAALALVRLGADDSLSQIDKLHV	60		
QY	61	NTASGYNSSNSGQSLQKRVFSDINASHKDDVSLVNGLFKAYVGFHKDYECAEK	120		
DB	61	NTASGYNSSNSGQSLQKRVFSDINASHKDDVSLVNGLFKAYVGFHKDYECAEK	120		
QY	121	LYDAKVRVDFPTNLEDTRRNINKVENEETHGKTKNYIGEGGSSSAVMVLVNAVYFKG	180		
DB	121	LYDAKVRVDFPTNLEDTRRNINKVENEETHGKTKNYIGEGGSSSAVMVLVNAVYFKG	180		
QY	181	WOSAFKSETINCHFKSPKSGKAVAMHDERKRLNSVIEDPSKILRLRYNGGINNYVL	240		
DB	181	WOSAFKSETINCHFKSPKSGKAVAMHDERKRLNSVIEDPSKILRLRYNGGINNYVL	240		

```

QY      241 LPENDLSEIENKLTQONLMEWTNPRMTSKYVEVFPQRIKENYEMKOYLALGLKIDIF 300
      |||
Db      241 LPENDLSEIENKLTQONLMEWTNPRMTSKYVEVFPQRIKENYEMKOYLALGLKIDIF 300

QY      301 DESKADLSGIASGGRLYISRMHKSXYIEVTEGTEATAATGSNIYVKOLPOSTLFRADHP 360
      |||
Db      301 DESKADLSGIASGGRLYISRMHKSXYIEVTEGTEATAATGSNIYVKOLPOSTLFRADHP 360

QY      361 FLFVIRKDDIILFSGKVSCP 380
      |||
Db      361 FLFVIRKDDIILFSGKVSCP 380

```

## RESULT 2

```

US-10-126-052A-263
; Sequence 263, Application US/10126052A
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Murray, Richard
; APPLICANT: Bos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of diagnosis of Lung Cancer, Compositions and
; FILE REFERENCE: 018501-001300US
; CURRENT APPLICATION NUMBER: US/10/126, 052A
; PRIOR FILING DATE: 2002-04-18
; PRIOR APPLICATION NUMBER: US 60/284,770
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: US 60/290,492
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: US 60/339,245
; PRIOR FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/334,370
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/372,246
; PRIOR FILING DATE: 2002-04-12
; NUMBER OF SEQ ID NOS: 691
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 263
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-126-052A-263

```

```

Query Match      100.0%; Score 1959; DB 12; Length 380;
Best Local Similarity 100.0%; Pred. No. 1.4e-181;
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 MASLAAANAEFCNLFREDDNOGNGNVFSSLSLFAALALVRLGAODDSLQIDKILHV 60
      |||
Db      1 MASLAAANAEFCNLFREDDNOGNGNVFSSLSLFAALALVRLGAODDSLQIDKILHV 60

QY      61 NTASGYGNSNSQSGIOSQLKRVFSDINASHKDYDIYNGLFAEKVYGFHHDYTCAR 120
      |||
Db      61 NTASGYGNSNSQSGIOSQLKRVFSDINASHKDYDIYNGLFAEKVYGFHHDYTCAR 120

QY      121 LVDKVERVDFNHLIEDTRRNINKWVENETHGKIKRVIGEGGSSSAVAVLVNAVYFKG 180
      |||
Db      121 LVDKVERVDFNHLIEDTRRNINKWVENETHGKIKRVIGEGGSSSAVAVLVNAVYFKG 180

QY      181 MOSAFKSETINCHFSKPCSGKAVAMMOERKFNLSYEDPSMKLLEIRYNGIMYVL 240
      |||
Db      181 MOSAFKSETINCHFSKPCSGKAVAMMOERKFNLSYEDPSMKLLEIRYNGIMYVL 240

QY      241 LPENDLSEIENKLTQONLMEWTNPRMTSKYVEVFPQRIKENYEMKOYLALGLKIDIF 300
      |||
Db      241 LPENDLSEIENKLTQONLMEWTNPRMTSKYVEVFPQRIKENYEMKOYLALGLKIDIF 300

QY      301 DESKADLSGIASGGRLYISRMHKSXYIEVTEGTEATAATGSNIYVKOLPOSTLFRADHP 360
      |||
Db      301 DESKADLSGIASGGRLYISRMHKSXYIEVTEGTEATAATGSNIYVKOLPOSTLFRADHP 360

```

```

QY      361 FLFVIRKDDIILFSGKVSCP 380
      |||
Db      361 FLFVIRKDDIILFSGKVSCP 380

```

## RESULT 3

```

US-60-443-566-3006
; Sequence 3006, Application US/60443566
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: BEGOVICH, Ann
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001447
; CURRENT APPLICATION NUMBER: US/60/443,566
; PRIOR FILING DATE: 2003-01-30
; NUMBER OF SEQ ID NOS: 25102
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3006
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-443-566-3006

```

```

Query Match      100.0%; Score 1959; DB 14; Length 380;
Best Local Similarity 100.0%; Pred. No. 1.4e-181;
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 MASLAAANAEFCNLFREDDNOGNGNVFSSLSLFAALALVRLGAODDSLQIDKILHV 60
      |||
Db      1 MASLAAANAEFCNLFREDDNOGNGNVFSSLSLFAALALVRLGAODDSLQIDKILHV 60

QY      61 NTASGYGNSNSQSGIOSQLKRVFSDINASHKDYDIYNGLFAEKVYGFHHDYTCAR 120
      |||
Db      61 NTASGYGNSNSQSGIOSQLKRVFSDINASHKDYDIYNGLFAEKVYGFHHDYTCAR 120

QY      121 LVDKVERVDFNHLIEDTRRNINKWVENETHGKIKRVIGEGGSSSAVAVLVNAVYFKG 180
      |||
Db      121 LVDKVERVDFNHLIEDTRRNINKWVENETHGKIKRVIGEGGSSSAVAVLVNAVYFKG 180

QY      181 MOSAFKSETINCHFSKPCSGKAVAMMOERKFNLSYEDPSMKLLEIRYNGIMYVL 240
      |||
Db      181 MOSAFKSETINCHFSKPCSGKAVAMMOERKFNLSYEDPSMKLLEIRYNGIMYVL 240

QY      241 LPENDLSEIENKLTQONLMEWTNPRMTSKYVEVFPQRIKENYEMKOYLALGLKIDIF 300
      |||
Db      241 LPENDLSEIENKLTQONLMEWTNPRMTSKYVEVFPQRIKENYEMKOYLALGLKIDIF 300

QY      301 DESKADLSGIASGGRLYISRMHKSXYIEVTEGTEATAATGSNIYVKOLPOSTLFRADHP 360
      |||
Db      301 DESKADLSGIASGGRLYISRMHKSXYIEVTEGTEATAATGSNIYVKOLPOSTLFRADHP 360

```

## RESULT 4

```

US-60-455-444-4692
; Sequence 4692, Application US/60455444
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: BEGOVICH, Ann
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001455
; CURRENT APPLICATION NUMBER: US/60/455,444
; PRIOR FILING DATE: 2003-03-18
; NUMBER OF SEQ ID NOS: 50986
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4692
; LENGTH: 380
; TYPE: PRT

```

ORGANISM: Homo sapiens  
US-60-465-444-4692

Query Match 100.0%; Score 1959; DB 14; Length 380;  
Best Local Similarity 100.0%; Pred. No. 1.4e-181;  
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASLAANAEEFCNLFREMDNOGNGVNFSSLSLFAALALVRLGADDSLSQIDKLHV 60  
DB 1 MASLAANAEEFCNLFREMDNOGNGVNFSSLSLFAALALVRLGADDSLSQIDKLHV 60  
QY 61 NTASGYGNSNSGSGLOSLKRVFSDINASHKDYDLSIVNGLEAEKYYGFHKDYIECAEK 120  
DB 61 NTASGYGNSNSGSGLOSLKRVFSDINASHKDYDLSIVNGLEAEKYYGFHKDYIECAEK 120  
QY 121 LYDAKVERVDFTNHLLEDTNRINKWVENETHGKIKVYIGGGISSAVMVLVNAVYFKG 180  
DB 121 LYDAKVERVDFTNHLLEDTNRINKWVENETHGKIKVYIGGGISSAVMVLVNAVYFKG 180  
QY 181 WQSAFTKSETINCHFKSPKCSGKAVAMHOREKFNLSVIDPMSKILELRNGGINNYVL 240  
DB 181 WQSAFTKSETINCHFKSPKCSGKAVAMHOREKFNLSVIDPMSKILELRNGGINNYVL 240  
QY 241 LPENDLSEIBNKLTFOULMEWTNPRRMTSKYVEVFPOFKIEKNYEKKOYLRLGLKIDIF 300  
DB 241 LPENDLSEIBNKLTFOULMEWTNPRRMTSKYVEVFPOFKIEKNYEKKOYLRLGLKIDIF 300  
QY 301 DESKADLSGIASGRLYISRMHMKSYIEVEEGTEATAAGSNIVEKQLPOSTLFRADHP 360  
DB 301 DESKADLSGIASGRLYISRMHMKSYIEVEEGTEATAAGSNIVEKQLPOSTLFRADHP 360  
QY 361 FLFVIRKDDIILFSGKAVSCP 380  
DB 361 FLFVIRKDDIILFSGKAVSCP 380

## RESULT 5

US-60-465-241-4692

Sequence 4692, Application US/60465241

GENERAL INFORMATION:

APPLICANT: CARGILL, Michele

APPLICANT: BEGOVICH, Ann

TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH

TITLE OF INVENTION: RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF

FILE REFERENCE: C0001468

CURRENT APPLICATION NUMBER: US/60/465,241

NUMBER OF SEQ ID NOS: 258418

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 4692

LENGTH: 380

TYPE: PRT

ORGANISM: Homo sapiens

US-60-465-241-4692

Query Match 100.0%; Score 1959; DB 14; Length 380;  
Best Local Similarity 100.0%; Pred. No. 1.4e-181;  
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASLAANAEEFCNLFREMDNOGNGVNFSSLSLFAALALVRLGADDSLSQIDKLHV 60  
DB 1 MASLAANAEEFCNLFREMDNOGNGVNFSSLSLFAALALVRLGADDSLSQIDKLHV 60  
QY 61 NTASGYGNSNSGSGLOSLKRVFSDINASHKDYDLSIVNGLEAEKYYGFHKDYIECAEK 120  
DB 61 NTASGYGNSNSGSGLOSLKRVFSDINASHKDYDLSIVNGLEAEKYYGFHKDYIECAEK 120  
QY 121 LYDAKVERVDFTNHLLEDTNRINKWVENETHGKIKVYIGGGISSAVMVLVNAVYFKG 180  
DB 121 LYDAKVERVDFTNHLLEDTNRINKWVENETHGKIKVYIGGGISSAVMVLVNAVYFKG 180  
QY 181 WQSAFTKSETINCHFKSPKCSGKAVAMHOREKFNLSVIDPMSKILELRNGGINNYVL 240  
DB 181 WQSAFTKSETINCHFKSPKCSGKAVAMHOREKFNLSVIDPMSKILELRNGGINNYVL 240

DB 181 WQSAFTKSETINCHFKSPKCSGKAVAMHOREKFNLSVIDPMSKILELRNGGINNYVL 240  
QY 241 LPENDLSEIBNKLTFOULMEWTNPRRMTSKYVEVFPOFKIEKNYEKKOYLRLGLKIDIF 300  
DB 241 LPENDLSEIBNKLTFOULMEWTNPRRMTSKYVEVFPOFKIEKNYEKKOYLRLGLKIDIF 300  
QY 301 DESKADLSGIASGRLYISRMHMKSYIEVEEGTEATAAGSNIVEKQLPOSTLFRADHP 360  
DB 301 DESKADLSGIASGRLYISRMHMKSYIEVEEGTEATAAGSNIVEKQLPOSTLFRADHP 360  
QY 361 FLFVIRKDDIILFSGKAVSCP 380  
DB 361 FLFVIRKDDIILFSGKAVSCP 380

## RESULT 6

US-09-936-883C-19

Sequence 19, Application US/09936883C

GENERAL INFORMATION:

APPLICANT: MIYATA, Toshio

TITLE OF INVENTION: A Method for Detecting Megsin Protein and Use

FILE REFERENCE: F2-101DIPCT

CURRENT APPLICATION NUMBER: US/09/936, 883C

CURRENT FILING DATE: 2001-12-21

PRIOR APPLICATION NUMBER: JP 1999-75305

PRIOR FILING DATE: 1999-03-19

PRIOR APPLICATION NUMBER: JP 1999-306623

NUMBER OF SEQ ID NOS: 21

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 19

LENGTH: 380

TYPE: PRT

ORGANISM: Rattus norvegicus

FEATURE:

NAME/KEY: misc\_feature

LOCATION: 51, 94

OTHER INFORMATION: Xaa is unknown.

US-09-936-883C-19

Query Match 75.4%; Score 1477; DB 10; Length 380;  
Best Local Similarity 73.4%; Pred. No. 9e-135;  
Matches 279; Conservative 56; Mismatches 45; Indels 0; Gaps 0;

QY 1 MASLAANAEEFCNLFREMDNOGNGVNFSSLSLFAALALVRLGADDSLSQIDKLHV 60  
DB 1 MASLAANAEEFCNLFREMDNOGNGVNFSSLSLFAALALVRLGADDSLSQIDKLHV 60  
QY 61 NTASGYGNSNSGSGLOSLKRVFSDINASHKDYDLSIVNGLEAEKYYGFHKDYIECAEK 120  
DB 61 NTASGYGNSNSGSGLOSLKRVFSDINASHKDYDLSIVNGLEAEKYYGFHKDYIECAEK 120  
QY 121 LYDAKVERVDFTNHLLEDTNRINKWVENETHGKIKVYIGGGISSAVMVLVNAVYFKG 180  
DB 121 LYDAKVERVDFTNHLLEDTNRINKWVENETHGKIKVYIGGGISSAVMVLVNAVYFKG 180  
QY 181 WQSAFTKSETINCHFKSPKCSGKAVAMHOREKFNLSVIDPMSKILELRNGGINNYVL 240  
DB 181 WQSAFTKSETINCHFKSPKCSGKAVAMHOREKFNLSVIDPMSKILELRNGGINNYVL 240  
QY 241 LPENDLSEIBNKLTFOULMEWTNPRRMTSKYVEVFPOFKIEKNYEKKOYLRLGLKIDIF 300  
DB 241 LPENDLSEIBNKLTFOULMEWTNPRRMTSKYVEVFPOFKIEKNYEKKOYLRLGLKIDIF 300  
QY 301 DESKADLSGIASGRLYISRMHMKSYIEVEEGTEATAAGSNIVEKQLPOSTLFRADHP 360  
DB 301 VESRADLSGIASGRLYISRMHMKSYIEVEEGTEATAAGSNIVEKQLPOSTLFRADHP 360  
QY 361 FLFVIRKDDIILFSGKAVSCP 380  
DB 361 FLFVIRKDDIILFSGKAVSCP 380





[illegible]

```

Query Match Similarity      42.0%; Score 832; DB 14; length 372;
Best Local Similarity      43.1%; Pred. No. 3.6e-71;
Matches 166; Conservative  75; Mismatches 126; Indels 18; Gaps 4

QY      1 MASLAAAAAEFCFLNLFREMDNOCNGNVAFFSSLSLFAALATATRLGADDSLSQIDKLHY 60
DB      1 MGSLSIAVVEFCLODFKELNSNIGDNIFFSSLSLLYALSMVLLGARGETEQLKYLHF 60
QY      61 NTASGYGSSNSQSGLSQSLKRVESDINASHKDYLSVNGLFAEKVYGFHKDYIECAEK 120
DB      61 -----SHTYDSLKPGFEDSKPDSNCTLSIANRLYGTGTMAFHQOYLSCSEK 107
QY      121 LYDAKVERVDFTNLHEDTRNINKKVENETHGKIKNNVGEGSSAAVVLVNNAYFFCK 180
DB      108 WYQAQLQVDEQOSTEETRKRTINAVENKTNCKVNLGKSTIDSSVAVLNALYFFKQ 167
QY      181 MOSARTSETINCHFKSPKSCGSAVAAVMOEKFNLSTIEDPSMKLELRY--NGGINVY 239
DB      168 WOKNKQVREYASPPQLSEGGKNVYEMMYQISTCFKLAFAKFEQOMOVLELPYNNKLSMTI 227
QY      240 LLPER--DLSEIEKNLFTFONLMEETNPRIKTSKYVEVFPQKIKKNYEMKQYLALGK 297
DB      228 LLPVAINLKOIEKRLNGSTFHEWYSSSMKMEREVNHLPRKLEIKELYNELLSLSGYT 287
QY      298 DIFDESKADLGSIASGRLYISRMHKSXYIEVEEGTEATAATGASNIYEKQLPOSTLFRA 357
DB      288 DLFNQVKADLSGMSPTKGLYLSKALIHKSXYLDVSEGTEAATAATGSIANKSLPRAQFRA 347
QY      358 DHPLEFVIR--KDIILFSGKAVSCP 380
DB      348 NHPLEFIRHTHTNTILFCGKLASP 372

RESULT 11
US-10-198-070-62
; Sequence 62, Application US/10198070
; GENERAL INFORMATION:
; APPLICANT: AVERBACK, PAUL
; APPLICANT: GEMMELL, JACK
; TITLE OF INVENTION: PEPTIDES EFFECTIVE IN THE TREATMENT OF TUMORS AND OTHER

```

```

: TITLE OF INVENTION: CONDITIONS REQUIRING THE REMOVAL OR DESTRUCTION OF
: FILE REFERENCE: 59003,000008
: CURRENT APPLICATION NUMBER: US/10/198,070
: PRIORITY FILING DATE: 2002-07-19
: PRIOR APPLICATION NUMBER: 60/306,161
: PRIOR FILING DATE: 2001-07-19
: PRIOR APPLICATION NUMBER: 60/306,150
: PRIOR FILING DATE: 2001-07-19
: PRIOR APPLICATION NUMBER: 60/331,477
: PRIOR FILING DATE: 2001-11-16
: NUMBER OF SEQ ID NOS: 125
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 62
: LENGTH: 392
: TYPE: prt
: ORGANISM: Homo sapiens
: US-10-198-070-62

```

```

Query Match          42.0%: Score 822; DB 12; Length 392;
Best Local Similarity 43.1%: Pred. No. 3.9e-71;
Matches 169; Conservative 80; Mismatches 131; Indels 12; Gaps 6

QY      1 MASLAAAAAECEFNLFREMDNNOGNGANFFSSLISFLAALATVLGAODSLISQIRLLHY 60
DB      1 MGSLSTANVECCLDVFKRELNSNIGNDITFSSLISLYALSMVLLGARGETAQLEKLVLF 60
QY      61 -----NTASGYGNSSN-SQSG-LOSOLKRVESDINASHKXDLSIVNGLFAEKVYGFHKD 113
DB      61 SHTVDSLKPGEFSDKSCQAGRIHSEFGVEFSQINOPDSNCLTSLIANRLYGRKTMAFHOQ 120
QY      114 YTECEAKLYDAKVENVDVTNHLIEDTRINKKVENVEETHGKIKVNGEGGSISSAMVVLVN 173
DB      121 YLSCSEKGYQARLDQVDFEOSTEERFKTINAVENKTNGKANLEGGSTIDPSSVMVLVN 180
QY      174 AVYFEGKQSAFETKSETINTCHFKSPSCGKAAVAMHOERKFNLSYIEDPSMKILELR-N 232
DB      181 TTFYGGQGNKRFQVAYELKSPFQLSBGNKYVEMKQIGTFFLAVKBERQMYLELPLVN 240
QY      223 GGINMYVLLPEN--DLSEIENKLTFFQNLMEWTNPRRPTSKYVEVFPQKIEKNYEMQY 290
DB      241 NKLSMIILLPVGIANLQKIQEQLNSGTFHEWMTSSNMMEEREVEVLPKREIKKIELNSL 300
QY      291 LRALGLKQIFDESKADSLGSIASGRLYISRMHKKYIEETEGTATATATSGNIYEKQLP 350
DB      301 LKPLGVTLDFQVQKADLSGMSPTKGLYLSKAIHKSLYLDVSEGTAAATGDSIAVKSLLP 360
QY      351 OSTFRADHPFLFVIR--KDDITLPSGKSCP 380
DB      361 MRAQFKANPFLFETIRHTHTNTILFCGLASP 392

RESULT 12
US-10-198-070-67
; Sequence 67, Application US/10198070
; GENERAL INFORMATION:
; APPLICANT: AVERBACK, PAUL
; APPLICANT: GEMMELL, JACK
; TITLE OF INVENTION: PEPTIDES EFFECTIVE IN THE TREATMENT OF TUMORS AND OTHER
; TITLE OF INVENTION: CONDITIONS REQUIRING THE REMOVAL OR DESTRUCTION OF
; TITLE OF INVENTION: CELLS
; FILE REFERENCE: 59003.000008
; CURRENT APPLICATION NUMBER: US/10/198,070
; CURRENT FILING DATE: 2002-07-19
; PRIOR APPLICATION NUMBER: 60/306,161
; PRIOR FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: 60/306,150
; PRIOR FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: 60/331,477
; PRIOR FILING DATE: 2001-11-16
; NUMBER OF SEQ ID NOS: 125
; SOFTWARE: Patentlin Ver. 2.1
; SEQ ID NO 67

```

LENGTH: 392  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 us-10-198-070-67

Query Match 41.9%; Score 820; DB 12; Length 392;

Best Local Similarity 43.1%; Pred. No. 6.2e-71;

Matches 169; Conservative 80; Mismatches 131; Indels 12; Gaps 6;

QY 1 MASLAANAFECNLFREDDNGNGNVEFSSLSLEPAALAVRLGAODDSLSQIDKLHY 60  
 DB 1 MGSLSLANVEFCIDVKEKLSNNIGDNIFSSLSLYALSMVLLGARGETAEQLEKVLHF 60  
 QY 61 -----NTASGYGNSSN-SOSG-LQSOLKRVFSDINASHKDYLSTYNGLEFAEKVYGFHKD 113  
 DB 61 SHTVDSLKPFGKDSPCSOAGRIHSEGFVEFSQINOPDSNCTLSIANRLYGTMAFHOQ 120  
 QY 114 YIECAKLDKAKVERDFTNHLNEDTRNINKWYENETHGKIKVIGEGGSSAVAVLVN 173  
 DB 121 YLSCSEKMYQARLQTYDFEOSTETETRKMINAWENTNGKAVANLFKSTIDPSSVAVLVN 180  
 QY 174 AVFEKGMOSAFETSETINCHFKSPKCSGKAVAMHOERKNLSVIEDPSMKITELRY-N 232  
 DB 181 TIYFKGQRNKFVRETVSPFOLSEGNVTVMQIGTFKLAFAKEPOMVLELRYN 240  
 QY 233 GGIMMYVLLPEN--DLSEIENKLTFLONLMEWTNPRMTSKYVEVFFPOFKIEKNYEMKQY 290  
 DB 241 NKLSMITLLPVGIANKQIEKQLNSGTFHEWTSNNMEREVEVHLPRFLEIKELYNSL 300  
 QY 291 LRLALGLDIFDESKADLSGASGRLYISRMHKSITYEVEETGTAATGSIYVOKLP 350  
 DB 301 LKPLGVTDLFNOVKADLSGMSPTKGLYLSKAIHKSIVLDVSEGTAAATGDSIAVKSIP 360  
 QY 351 OSTLFRAHDHFLFVIR--KDDIILFSGKVSCEP 380  
 DB 361 MRAQFKAHPFLFFIRHTHTNTILFCGKLASP 392

RESULT 13

us-10-198-070-70  
 Sequence 70, Application US/10198070

GENERAL INFORMATION:  
 APPLICANT: AVERBACK, PAUL  
 APPLICANT: GEMMELL, JACK  
 TITLE OF INVENTION: PEPTIDES EFFECTIVE IN THE TREATMENT OF TUMORS AND OTHER  
 TITLE OF INVENTION: CONDITIONS REQUIRING THE REMOVAL OR DESTRUCTION OF  
 TITLE OF INVENTION: CELLS  
 FILE REFERENCE: 59003.000008  
 CURRENT APPLICATION NUMBER: US/10/198, 070  
 CURRENT FILING DATE: 2002-07-19  
 PRIOR APPLICATION NUMBER: 60/306, 161  
 PRIOR FILING DATE: 2001-07-19  
 PRIOR APPLICATION NUMBER: 60/306, 150  
 PRIOR FILING DATE: 2001-07-19  
 PRIOR APPLICATION NUMBER: 60/331, 477  
 PRIOR FILING DATE: 2001-11-16  
 NUMBER OF SEQ ID NOS: 125  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 70  
 LENGTH: 392  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 us-10-198-070-70

Query Match 41.8%; Score 819; DB 12; Length 392;

Best Local Similarity 43.1%; Pred. No. 7.7e-71;

Matches 169; Conservative 80; Mismatches 131; Indels 12; Gaps 6;

QY 1 MASLAANAFECNLFREDDNGNGNVEFSSLSLEPAALAVRLGAODDSLSQIDKLHY 60  
 DB 1 MGSLSLANVEFCIDVKEKLSNNIGDNIFSSLSLYALSMVLLGARGETAEQLEKVLHF 60  
 QY 61 -----NTASGYGNSSN-SOSG-LQSOLKRVFSDINASHKDYLSTYNGLEFAEKVYGFHKD 113

DB 61 SHTVDSLKPFGKDSPCSOAGRIHSEGFVEFSQINOPDSNCTLSIANRLYGTMAFHOQ 120  
 QY 114 YIECAKLDKAKVERDFTNHLNEDTRNINKWYENETHGKIKVIGEGGSSAVAVLVN 173  
 DB 121 YLSCSEKMYQARLQTYDFEOSTETETRKMINAWENTNGKAVANLFKSTIDPSSVAVLVN 180  
 QY 174 AVFEKGMOSAFETSETINCHFKSPKCSGKAVAMHOERKNLSVIEDPSMKITELRY-N 232  
 DB 181 TIYFKGQRNKFVRETVSPFOLSEGNVTVMQIGTFKLAFAKEPOMVLELRYN 240  
 QY 233 GGIMMYVLLPEN--DLSEIENKLTFLONLMEWTNPRMTSKYVEVFFPOFKIEKNYEMKQY 290  
 DB 241 NKLSMITLLPVGIANKQIEKQLNSGTFHEWTSNNMEREVEVHLPRFLEIKELYNSL 300  
 QY 291 LRLALGLDIFDESKADLSGASGRLYISRMHKSITYEVEETGTAATGSIYVOKLP 350  
 DB 301 LKPLGVTDLFNOVKADLSGMSPTKGLYLSKAIHKSIVLDVSEGTAAATGDSIAVKSIP 360  
 QY 351 OSTLFRAHDHFLFVIR--KDDIILFSGKVSCEP 380  
 DB 361 MRAQFKAHPFLFFIRHTHTNTILFCGKLASP 392

RESULT 14

us-10-294-025-946  
 Sequence 946, Application US/10294025

GENERAL INFORMATION:  
 APPLICANT: Xu, Jiangchun  
 APPLICANT: Scolik, John A.  
 APPLICANT: Kalos, Michael D.  
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
 TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER  
 FILE REFERENCE: 210121.427C29  
 CURRENT APPLICATION NUMBER: US/10/294, 025  
 CURRENT FILING DATE: 2002-11-12  
 NUMBER OF SEQ ID NOS: 1038  
 SOFTWARE: FastSeq for Windows Version 3.0  
 SEQ ID NO 946  
 LENGTH: 379  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 us-10-294-025-946

Query Match 40.7%; Score 796.5; DB 12; Length 379;

Best Local Similarity 42.6%; Pred. No. 1.1e-68;

Matches 168; Conservative 77; Mismatches 120; Indels 29; Gaps 8;

QY 1 MASLAANAFECNLFREDDNGNGNVEFSSLSLEPAALAVRLGAODDSLSQIDKLHY 60  
 DB 1 MGSLSLANVEFCIDVKEKLSNNIGDNIFSSLSLYALSMVLLGARGETAEQLEKVLHF 60  
 QY 61 -----NTASGYGNSSN-SOSG-LQSOLKRVFSDINASHKDYLSTYNGLEFAEKVYGFHKD 113  
 DB 61 SHTVDSLKPFGKDSPCSOAGRIHSEGFVEFSQINOPDSNCTLSIANRLYGTMAFHOQ 120  
 QY 114 YIECAKLDKAKVERDFTNHLNEDTRNINKWYENETHGKIKVIGEGGSSAVAVLVN 173  
 DB 121 YLSCSEKMYQARLQTYDFEOSTETETRKMINAWENTNGKAVANLFKSTIDPSSVAVLVN 180  
 QY 174 AVFEKGMOSAFETSETINCHFKSPKCSGK--AVAMHOERKNLSVIEDPSMKITELRY 231  
 DB 181 TIYFKGQRNKF-----QGKNVYEMMYOIGTFKLAFAKEPOMVLELRY 225  
 QY 232 -NGIMMYVLLPEN--DLSEIENKLTFLONLMEWTNPRMTSKYVEVFFPOFKIEKNYEMK 288  
 DB 226 VNNKLSMITLLPVGIANKQIEKQLNSGTFHEWTSNNMEREVEVHLPRFLEIKELYNSL 285  
 QY 289 QYRALGLDIFDESKADLSGASGRLYISRMHKSITYEVEETGTAATGSIYVOKLP 348  
 DB 286 SLKPLGVTDLFNOVKADLSGMSPTKGLYLSKAIHKSIVLDVSEGTAAATGDSIAVKS 345  
 QY 349 LPSTLFRAHDHFLFVIR--KDDIILFSGKVSCEP 380



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:42:16 ; Search time 5.16424 Seconds  
(without alignments)  
232.223 Million cell updates/sec

Title: US-10-091-442-1  
Perfect score: 52  
Sequence: 1 SETINCHFK 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

```
Searched:      908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters:  908470
```

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
A\_Genseq\_101002 : \*

1:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA1980.DAT :	*
2:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA1981.DAT :	*
3:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA1982.DAT :	*
4:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA1983.DAT :	*
5:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA1984.DAT :	*
6:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA1985.DAT :	*
7:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA1986.DAT :	*
8:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA1987.DAT :	*
9:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA1988.DAT :	*
10:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA1989.DAT :	*
11:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA1990.DAT :	*
12:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA1991.DAT :	*
13:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA1992.DAT :	*
14:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA1993.DAT :	*
15:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA1994.DAT :	*
16:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA1995.DAT :	*
17:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA1996.DAT :	*
18:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA1997.DAT :	*
19:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA1998.DAT :	*
20:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA1999.DAT :	*
21:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA2000.DAT :	*
22:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA2001.DAT :	*
23:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA2002.DAT :	*

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	52	100.0	360	15	AA048379	Human megakaryocyte
2	52	100.0	360	20	AA008254	Human megasin protein
3	52	100.0	360	21	AA024142	Human megasin prote
4	52	100.0	360	22	AA083075	Human megasin prote
5	39	75.0	360	20	AA008255	Human megasin protei
6	39	75.0	360	21	AA024150	Rat megasin protein
7	39	75.0	360	22	AA064286	Rat megasin protein
8	39	75.0	360	22	AA083076	Rat megasin proteina
9	38	73.1	199	22	AA062342	Drosophila melanog
10	38	73.1	208	22	AA060620	Drosophila melanog

11	37	71.2	9	15	AAR57107	Human megakaryocyt
12	37	71.2	320	16	AAR79450	Rat A3 adenosine r
13	37	71.2	1030	23	ABP35638	Fungal ZBC protein
14	36	69.2	64	22	AAO09905	Human polypeptide
15	36	69.2	863	23	ABP35700	Fungal ZBC protein
16	36	69.2	1354	22	ABB62359	Drosophila melanog
17	36	69.2	2176	22	ABBB62629	Drosophila melanog
18	35	67.3	90	22	AAU87605	Novel central nerv
19	35	67.3	90	22	ABBI7012	Human nervous syst
20	35	67.3	90	22	ABBI0308	Human CDNA SBO ID
21	35	67.3	90	22	AAU19916	Novel human calciu
22	35	67.3	856	9	AAAP80803	Sequence of env pr
23	35	67.3	858	9	AAAP80806	Sequence of env pr
24	35	67.3	858	9	AAAP81779	Sequence encoded b
25	35	67.3	858	13	AAAR20604	ROD HIV-2 Env prot
26	35	67.3	858	21	AAV51984	HIV-2 ROD isolate
27	35	67.3	890	23	AAU96798	Human immunodefici
28	35	67.3	890	23	AAU96799	Human immunodefici
29	35	67.3	890	23	AAU96804	Human immunodefici
30	35	67.3	891	9	AAAP2677	Human immunodefici
31	35	67.3	891	23	AAU96794	Human immunodefici
32	35	67.3	974	21	AAV45000	Human calcium sens
33	34	65.4	60	22	ABBI5580	Human nervous syst
34	34	65.4	70	20	AAV36023	Extended human sec
35	34	65.4	105	22	AAU14892	Novel bone marrow
36	34	65.4	120	23	ABGG6683	Human novel polypep
37	34	65.4	123	23	ABGG6682	Human novel polypep
38	34	65.4	159	22	ABGI6246	Novel human diagn
39	34	65.4	185	13	AAAR2263	Coilagen type prob
40	34	65.4	195	20	AAV34440	Porphyrinomas ging
41	34	65.4	195	20	AAV99071	Partial human PIGR1
42	34	65.4	201	20	AAV34319	Porphyrinomas ging
43	34	65.4	201	20	AAV993327	Human PIGR-1, Hom
44	34	65.4	201	22	AAU293327	Human PRO polypep
45	34	65.4	201	22	AAAB31568	Amino acid sequenc

## ALIGNMENTS

RESULT 1	
AAR48379	
ID	AAR48379 standard; Protein; 380 AA.
XX	
AC	AAR48379;
XX	
DT	16-AUG-1994 (first entry)
XX	
DE	Human megakaryocyte differentiation factor.
XX	
KW	Human megakaryocyte differentiation factor; MDF; thrombopoietin
KW	hematopoietic stimulating factor; thrombocytopoenta; platelet;
XX	bone marrow transplantation; cancer chemotherapy.
XX	
OS	Homo sapiens.
XX	
PN	EP583884-A.
XX	
PD	23-FEB-1994.
XX	
PF	19-JUL-1993; 93EP-0305654.
XX	
PR	17-JUL-1992; 92JP-0212305.
PR	04-MAR-1993; 93JP-0067339.
XX	
PA	(SUNR ) SUNTORY LTD.
PA	(TSUJ/) TSUJIMOTO M.
PI	Ishida N, Iwasa F, Kurihara T, Miura K, Nakazato H;
PI	Tsujimoto M, Tsuruoka N, Yamaguchi N, Yamachi K;
DR	WPI; 1994-058782/08.
DR	N-PSDB; AAQ56670.

```

XX New megakaryocyte differentiation factor - isolated from human
PT epidermoid carcinoma cells; used to treat conditions involving a
PT decrease in platelets
XX
PS Claim 7; Page 30-32; 47pp; English.
XX
CC Human MDF can be isolated from a culture of human epidermoid
CC carcinoma A431 cells in protein-free medium. The MDF stimulates
CC differentiation of megakaryocytes from myeloid cells in the presence
CC of IL-3. The MDF acts in vivo as a thrombopoietin making it useful
CC for treatment of diseases involving a decrease in platelet number
CC (esp. thrombocytopaenia) such as occurs in bone marrow
CC transplantation and in chemotherapy. MDF has mol.wt. 55-57kd by
CC SDS-PAGE and exhibits an isoelectric point of 6.0-7.0.
XX
SQ Sequence 380 AA;
Query Match 100.0%; Score 52; DB 15; Length 380;
Best Local Similarity 100.0%; Pred. No. 0.092;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SETINCHFK 9
Db 188 SETINCHFK 196

RESULT 2
AA08254
ID AA08254 standard; Protein: 380 AA.
XX
AC AA08254;
XX
DT 14-JUL-1999 (first entry)
XX
DE Human megsin protein.
XX
KW Megsin; mesangial cell; treatment; diagnosis; disease; Iga nephropathy;
KW human; rat; murine.
XX
OS Homo sapiens.
XX
PN MO9915652-A1.
XX
PD 01-APR-1999.
XX
PF 22-SEP-1998; 98WO-JP04269.
XX
PR 22-SEP-1997; 97JP-0275302.
XX
PA (KUROO/) KUROKAWA K.
PA (MIYA/) MIYATA T.
XX
PI Miyata T;
XX
DR WPI; 1999-276983/23.
DR N-PSDB; AAX56712.
XX
PT Megsin protein expressed specifically in mesangial cells
XX
PS Claim 1; Page 62-64; 100pp; Japanese.
XX
CC This invention describes the isolation of novel megsin nucleic acid and
CC proteins from human, rat and mouse tissue. This protein is expressed
CC specifically in mesangial cells. The products of the invention are
CC useful for the treatment and diagnosis of diseases involving mesangial
CC cells, such as Iga nephropathy.
XX
SQ Sequence 380 AA;
Query Match 100.0%; Score 52; DB 20; Length 380;
Best Local Similarity 100.0%; Pred. No. 0.092;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 SETINCHFK 9
Db 188 SETINCHFK 196

RESULT 3
AAB24142
ID AAB24142 standard; Protein: 380 AA.
XX
AC AAB24142;
XX
DT 30-JAN-2001 (first entry)
XX
DE Human megsin protein sequence SEQ ID NO:2.
XX
KW Megsin; mesangium-predominant gene; seipin regulated; nephropathy;
KW Iga; Immunoglobulin A; detection; renal function; renal disorder;
KW diagnosis; biological sample; blood; urine.
XX
OS Homo sapiens.
XX
PN WO200057189-A1.
XX
PD 28-SEP-2000.
XX
PF 17-MAR-2000; 2000WO-JP01646.
XX
PR 19-MAR-1999; 99JP-0075305.
PR 28-OCT-1999; 99JP-0306623.
XX
PA (KUROO/) KUROKAWA K.
PA (FUSO ) FUSO PHARM IND LTD.
PA (MIYA/) MIYATA T.
XX
PI Miyata T;
XX
DR WPI; 2000-611642/58.
DR N-PSDB; AAA99294.
XX
PT Evaluating renal function comprises assaying megsin protein in
PT biological sample
XX
PS Example 2; Page 66-69; 93pp; Japanese.
XX
CC The present invention describes a method for evaluating renal function.
CC The method comprises assaying megsin protein in biological sample. Also
CC described are: (1) use of a anti-megsin protein antibody for diagnosing
CC renal function; and (2) a kit for detecting megsin protein comprising:
CC (a) anti-megsin protein antibody attached to solid magnetic particles;
CC (b) direct or indirect fixing for the antibody to the particles; and
CC (c) a magnet. The process is useful for evaluating renal function and
CC diagnosing renal disorders by assaying megsin protein in biological
CC samples (preferably urine or blood). The process is reproducible and
CC gives accurate results. The present sequence represents the human megsin
CC protein, which is given in the exemplification of the present invention.
XX
SQ Sequence 380 AA;
Query Match 100.0%; Score 52; DB 21; Length 380;
Best Local Similarity 100.0%; Pred. No. 0.092;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SETINCHFK 9
Db 188 SETINCHFK 196

RESULT 4
AAB83075
ID AAB83075 standard; Protein: 380 AA.
AC AAB83075;

```

XX 10-JUL-2001 (first entry)  
XX Human mesgin protein.  
DE  
XX Human; mesgin: mesangial cell proliferative nephritis; nephrotropic;  
KW transgenic mouse; glomerular disease; animal model; drug screening.  
XX  
OS Homo sapiens.  
XX  
PN WO200124628-A1.  
XX  
PD 12-APR-2001.  
XX  
PE 06-OCT-2000; 2000WO-JP06988.  
XX  
PR 06-OCT-1999; 99JP-0285736.  
XX  
PA (KUROO/) KUROKAWA K.  
PA (MIYA/) MIYATA T.  
XX  
PI Miyata T;  
XX  
XX WPI: 2001-300136/31.  
DR N-PSDB: AAF82438.  
XX  
XX Mouse model for mesangial cell proliferative nephritis for development  
PT and screening of new treatments -  
XX  
PS Example 4; Page 44-46; 62pp; Japanese.  
XX  
CC The present sequence is human mesgin. The human mesgin coding  
CC sequence may be introduced into a mouse to produce an animal model of  
CC mesangial cell proliferative nephritis. The symptoms include  
CC enlargement of the mesangial base region, sedimentation of an immune  
CC complex and an increase in mesangial cells. The animal model is useful  
CC for analysing the pathology of chronic glomerular diseases and for  
CC screening compositions for prevention and treatment of the diseases.  
CC Highly uniform models can be made easily and in large numbers using  
CC this method.  
XX  
SQ Sequence 380 AA;  
XX  
OY 1 SETINCHFK 9  
DB 188 SETINCHFK 196

Query Match 100.0%; Score 52; DB 22; Length 380;  
Best Local Similarity 100.0%; Pred. NO. 0.092;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5  
AAV08255  
ID AAV08255 standard; Protein: 380 AA.  
XX  
XX  
AC AAY08255;  
XX  
DT 14-JUL-1999 (first entry)  
XX  
DE Rat mesgin protein.  
XX  
KW Mesgin; mesangial cell; treatment; diagnosis; disease; IGA nephropathy;  
KW human; rat; murine.  
XX  
OS Rattus rattus.  
XX  
PN WO9915652-A1.  
XX  
PD 01-APR-1999.  
XX  
PF 22-SEP-1998; 98WO-JP04269.  
XX

PR 22-SEP-1997; 97JP-0275302.  
XX  
XX (KUROO/) KUROKAWA K.  
PA (MIYA/) MIYATA T.  
XX  
XX Miyata T;  
XX  
DR WPI: 1999-276983/23.  
DR N-PSDB: AAX56712.  
XX  
XX Mesgin protein expressed specifically in mesangial cells  
PT  
PS Claim 1; Page 69-72; 100pp; Japanese.  
XX  
CC This invention describes the isolation of novel mesgin nucleic acid and  
CC proteins from human, rat and mouse tissue. This protein is expressed  
CC specifically in mesangial cells. The products of the invention are  
CC useful for the treatment and diagnosis of diseases involving mesangial  
CC cells, such as IGA nephropathy.  
XX  
SQ Sequence 380 AA;  
XX  
OY 1 SETINCHFK 9  
DB 188 SDTISCHFK 196

Query Match 75.0%; Score 39; DB 20; Length 380;  
Best Local Similarity 55.6%; Pred. NO. 26;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

RESULT 6  
AAB24150  
ID AAB24150 standard; Protein: 380 AA.  
XX  
XX AAB24150;  
AC  
XX  
DT 30-JAN-2001 (first entry)  
XX  
DE Rat mesgin protein sequence SEQ ID NO:19.  
XX  
XX  
KW Mesgin; mesangium-predominant gene; serpin regulated; nephropathy;  
KW IGA; immunoglobulin A; detection; renal function; renal disorder;  
KW diagnosis; biological sample; blood; urine.  
XX  
OS Rattus norvegicus.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 51  
FT FT /note= "unspecified"  
FT Misc-difference 94  
FT FT /note= "unspecified"  
XX  
PN WO200057189-A1.  
XX  
PD 28-SEP-2000.  
XX  
PF 17-MAR-2000; 2000WO-JP01646.  
XX  
XX 19-MAR-1999; 99JP-0075305.  
PR 28-OCT-1999; 99JP-0306623.  
XX  
PA (KUROO/) KUROKAWA K.  
PA (FUJO ) FUJO PHARM IND LTD.  
PA (MIYA/) MIYATA T.  
XX  
XX Miyata T;  
XX  
XX WPI: 2000-611642/58.  
DR N-PSDB: AAC55238.  
XX  
PT Evaluating renal function comprises assaying mesgin protein in  
PT biological sample -

xx	Example 2; Page 81-84; 93pp; Japanese.
p5	
cc	The present invention describes a method for evaluating renal function.
cc	The method comprises assaying mesgin protein in biological sample. Also
cc	described are: (1) use of an antimargin protein antibody for diagnosing
cc	renal function; and (2) a kit for detecting mesgin protein comprising:
cc	(a) antimargin protein antibody attached to solid magnetic particles;
cc	(b) direct or indirect fixing for the antibody to the particles; and
cc	(c) a magnet. The process is useful for evaluating renal function and
cc	diagnosing renal disorders by assaying mesgin protein in biological
cc	samples (preferably urine or blood). The process is reproducible and
cc	gives accurate results. The present sequence represents the rat mesgin
cc	protein, which is given in the exemplification of the present invention.
xx	
sq	Sequence 380 AA;
QY	Query Match Best Local Similarity 75.0%; Score 39; DB 21; Length 380; Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
D6	1 SETINCHFK 9 I:::    : 188 SDRLSCHF 196
RESULT 7	
ID	AAG64286 standard; Protein; 380 AA.
XX	AAG64286;
AC	
XX	21-SEP-2001 (first entry)
DT	
DE	Rat mesgin protein.
KW	Rat, mesgin; renal mesangial cell; mesangium proliferative nephritis.
XX	
OS	Rattus norvegicus.
PN	MO200148019-A1.
PD	05-JUL-2001.
PF	26-DEC-2000; 2000WO-JP09251.
PR	28-DEC-1999; 99JP-0373677.
XX	
PA	(KUROO/) KUROKAWA K.
PA	(MIYA/) MIYATA T.
PI	Miyata T;
PI	
DR	WPI: 2001-425651/45.
N-PSDB:	AAH48181.
XX	New antibody recognizing a partial sequence of rat mesgin protein for
PT	diagnosis of mesangium proliferative nephritis -
XX	
PS	Disclosure; Page 54-56; 63pp; Japanese.
XX	
CC	The present invention relates to a novel antibody which recognises a
CC	peptide consisting of residues 341-354 of rat mesgin protein. The present
CC	sequence is the protein sequence for rat mesgin, which was used in the
CC	present invention. Mesgin is highly expressed in renal mesangial cells
CC	and its level is elevated in mesangium proliferative nephritis. Assay of
CC	the serum or urine level using the antibody is therefore indicative of
CC	this type of disorder.
XX	
SQ	Sequence 380 AA;
Query Match	75.0%; Score 39; DB 22; Length 380;
Best Local Similarity	55.6%; Pred. No. 26;
Matches	5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
OY	1 SETINCHFK 9 I:::    : 188 SDRLSCHF 196
RESULT 8	
ID	AAB83076 standard; Protein; 380 AA.
XX	AAB83076;
AC	
XX	10-JUL-2001 (first entry)
DT	
DE	Rat mesgin protein.
KW	Rat, mesgin; mesangial cell proliferative nephritis; nephrotropic;
XX	transgenic mouse; glomerular disease; animal model; drug screening.
OS	Rattus norvegicus.
PN	WO200124628-A1.
PD	12-APR-2001.
PF	06-OCT-2000; 2000WO-JP06988.
PR	06-OCT-1999; 99JP-0285736.
XX	
PA	(KUROO/) KUROKAWA K.
PA	(MIYA/) MIYATA T.
PI	Miyata T;
PI	
DR	WPI: 2001-300136/31.
N-PSDB:	AAF82439.
XX	Mouse model for mesangial cell proliferative nephritis for development
PT	and screening of new treatments -
XX	
PS	Disclosure; Page 48-50; 62pp; Japanese.
XX	
CC	The present sequence is rat mesgin. The human mesgin coding
CC	sequence may be introduced into a mouse to produce an animal model of
CC	mesangial cell proliferative nephritis. The symptoms include
CC	enlargement of the mesangial base region, sedimentation of an immune
CC	complex and an increase in mesangial cells. The animal model is useful
CC	for analysing the pathology of chronic glomerular diseases and for
CC	screening compositions for prevention and treatment of the diseases.
CC	Highly uniform models can be made easily and in large numbers using
CC	this method.
XX	
SQ	Sequence 380 AA;
Query Match	75.0%; Score 39; DB 22; Length 380;
Best Local Similarity	55.6%; Pred. No. 26;
Matches	5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
OY	1 SETINCHFK 9 I:::    : 188 SDRLSCHF 196
RESULT 9	
ID	AAB62242 standard; Protein; 199 AA.
XX	AAB62242;
AC	
XX	26-MAR-2002 (first entry)
DT	
DE	Drosophila melanogaster polypeptide SEQ ID NO 13518.



XX Drosophila: developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 XX  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, LI PMD, Myers EW;  
 XX  
 DR WPI: 2001-656860/75.  
 DR N-PSDB; ABL06345.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 PS Disclosure: SEQ ID NO 13518; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 199 AA:  
 Query Match 73.1%; Score 38; DB 22; Length 199;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 3 TINCHF 8  
 Db 3 TINCHF 8  
 RESULT 10  
 ABB60820  
 ID ABB60820 standard; Protein; 208 AA.  
 XX  
 AC ABB60820;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 9252.  
 XX  
 KW Drosophila: developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 XX

PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, LI PMD, Myers EW;  
 XX  
 DR WPI: 2001-656860/75.  
 DR N-PSDB; ABL04923.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 PS Disclosure: SEQ ID NO 9252; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 208 AA:  
 Query Match 73.1%; Score 38; DB 22; Length 208;  
 Best Local Similarity 100.0%; Pred. No. 21;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 3 TINCHF 8  
 Db 3 TINCHF 8  
 RESULT 11  
 AAR57107  
 ID AAR57107 standard; peptide; 9 AA.  
 XX  
 AC AAR57107;  
 XX  
 DT 16-AUG-1994 (first entry)  
 XX  
 DE Human megakaryocyte differentiation factor peptide 1.  
 XX  
 KW Human megakaryocyte differentiation factor; MDF; thrombopoietin;  
 KW haematopoietic stimulating factor; thrombocytopoietin; platelet;  
 KW bone marrow transplantation; cancer chemotherapy.  
 XX  
 XX Homo sapiens.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FH Misc-difference 1 /note= "not determined"  
 FT Misc-difference 6 /note= "not determined"  
 FT  
 XX  
 PN EP583884-A.  
 XX  
 PD 23-FEB-1994.  
 XX  
 PF 19-JUL-1993; 93EP-0305654.  
 XX  
 PR 17-JUL-1992; 92JP-0212305.  
 PR 04-MAR-1993; 93JP-0067339.  
 XX  
 PA (SUNR ) SUNTORY LTD.  
 PA (TSUJ/) TSUJIMOTO M.  
 XX  
 PI Ishida N, Iwasa F, Kurihara T, Miura K, Nakazato H;

PI Tsujimoto M, Tsuruoka N, Yamaguchi N, Yamachi K;  
 XX WPI; 1994-058782/08.  
 XX  
 PT New megakaryocyte differentiation factor - isolated from human  
 PT epidermoid carcinoma cells, used to treat conditions involving a  
 PT decrease in platelets  
 XX  
 PS Claim 1; Page 18; 47pp; English.  
 XX  
 CC Human MDF (see AAR48379) can be isolated from a culture of human  
 CC epidermoid carcinoma A431 cells in protein-free medium. The MDF  
 CC stimulates differentiation of megakaryocytes from myeloid cells  
 CC in the presence of IL-3. The MDF acts in vivo as a thrombopoietin  
 CC making it useful for treatment of diseases involving a decrease  
 CC in platelet number (esp. thrombocytopenia) such as occurs in bone  
 CC marrow transplantation and in chemotherapy. MDF has mol wt  
 CC 55-57KD by SDS-PAGE, exhibits an isoelectric point of 6.0-7.0 and  
 CC contains an amino acid sequence comprising at least one of the  
 CC sequences AAR57107-R57115.  
 XX  
 SQ Sequence 9 AA;  
 XX  
 Query Match 71.2%; Score 37; DB 15; Length 9;  
 Best Local Similarity 87.3%; Pred. No. 7.8e+05;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 ETINCHFR 9  
 ||| |||  
 Db 2 ETINCHFR 9  
 2  
 RESULT 12  
 AAR79450  
 ID AAR79450 standard; Protein; 320 AA.  
 XX  
 AC AAR79450;  
 XX  
 DT 19-JAN-1996 (first entry)  
 XX  
 DE Rat A3 adenosine receptor.  
 XX  
 KW A3 adenosine receptor; screening; identification; agonist;  
 KW antagonist; drug; probe; detection; genetic disorder; RFLP;  
 KW restriction fragment length polymorphism.  
 XX  
 OS Rattus rattus.  
 XX  
 PN US5441883-A.  
 XX  
 PD 15-AUG-1995.  
 XX  
 PF 03-MAR-1992; 9205-0847563.  
 XX  
 PR 03-MAR-1992; 9205-0847563.  
 PR 02-AUG-1993; 9305-0101435.  
 XX  
 XX (UYOR-) UNIV OREGON HEALTH SCI.  
 PA  
 XX Clivelli O, Zhou Q;  
 PI  
 XX WPI; 1995-292492/38.  
 DR N-PSDB: AAO97312.  
 XX  
 PT DNA and protein sequence(s) for novel adenosine receptor, A3 - also  
 PT recombinant vector contg the DNA, DNA probe(s), and cell culture  
 PT transformed with the vector.  
 XX  
 PS Claim 1; Figure 1; 29pp; English.  
 XX  
 CC A3 adenosine receptors produced from cloned genes may be used to  
 CC screen compounds for A3 adenosine receptor activity, or for  
 CC determining the amount of adenosine agonist or antagonist drug in a

CC solution. The DNA or its fragments may also be used as probes to  
 CC determine tissue distribution of the receptors, to detect the  
 CC presence or absence of the gene or in RFLP to detect genetic  
 CC disorders.  
 XX  
 SQ Sequence 320 AA;  
 XX  
 Query Match 71.2%; Score 37; DB 16; Length 320;  
 Best Local Similarity 55.6%; Pred. No. 52;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 SETINCHFR 9  
 ||| |||  
 Db 163 SFTLCHFR 171  
 RESULT 13  
 ABP35638  
 ID ABP35638 standard; Protein; 1030 AA.  
 XX  
 AC ABP35638;  
 XX  
 DT 24-JUL-2002 (first entry)  
 XX  
 DE Fungal ZBC protein sequence #64.  
 XX  
 KW Secondary metabolite; fungus; ZBC gene; zinc binuclear cluster protein;  
 KW antibacterial; beta-lactam; anti-hypercholesterolaemic; lovastatin;  
 KW mevastatin; immunosuppressant; cyclosporin A; ergot alkaloid; ergotamine;  
 KW angiotensin inhibitor; ovalicin; glucan synthase inhibitor; gliotoxin;  
 KW fungal toxin; cell surface receptor; plant growth regulator; pigment;  
 KW insecticide; antineoplastic.  
 XX  
 OS Unidentified.  
 XX  
 PN WO200224865-A2.  
 XX  
 PD 28-MAR-2002.  
 XX  
 PF 19-SEP-2001; 2001WO-US29288.  
 XX  
 PR 19-SEP-2000; 2000US-233564P.  
 XX  
 PA (MICR-) MICROBIA INC.  
 XX  
 PI Holtzman D, Madden K, Maxon M, Sherman A;  
 XX  
 DR WPI; 2002-352005/38.  
 DR N-PSDB: ABN79827.  
 XX  
 PT New method for improving the production of a secondary metabolite e.g.  
 PT antineoplastic agent, ergot alkaloid from a fungus involves modulation  
 PT of the expression of at least one zinc binuclear cluster protein gene  
 PT  
 XX  
 XX Disclosure: SEQ ID 140; 49pp + sequence listing; English.  
 XX  
 CC The invention relates to improving the production of a secondary  
 CC metabolite by a fungus. This involves modulating the expression of at  
 CC least one ZBC (zinc binuclear cluster protein) gene in a manner to  
 CC improve the yield of the secondary metabolite. Methods of the invention  
 CC may be used for improving the production of the secondary metabolite e.g.  
 CC antibacterial (such as beta-lactam), an anti-hypercholesterolaemic (such  
 CC as lovastatin or mevastatin), an immunosuppressant (such as cyclosporin A),  
 CC an ergot alkaloid (such as ergotamine), an angiotensin inhibitor (such  
 CC as ovalicin), a glucan synthase inhibitor, gliotoxin family of compounds,  
 CC a fungal toxin, a modulator of cell surface receptor signalling, a plant  
 CC growth regulator, a pigment, an insecticide, or an antineoplastic  
 CC compound. The method results in a decrease in fermentor run-time, a  
 CC decrease in the size of the fermentor required for the production of  
 CC equivalent amounts of the secondary metabolite, or a decrease in the  
 CC biomass required for the production, which translates into decreased  
 CC waste that must be handled in downstream processing. The sequences given

CC In records ABP35575-ABP35722 represent ZBC proteins.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

CC Sequence 1030 AA;

Query Match 71.2%; Score 37; DB 23; Length 1030;  
Best Local Similarity 55.6%; Pred No. 1.8e+02;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 SETINCH 9

DB 689 SRVLCNHR 697

#### RESULT 14

AA009905 AAO09905 standard; Protein; 64 AA.

AC AAO09905;

DT 06-NOV-2001 (first entry)

DE Human polypeptide SEQ ID NO 23797.

KM Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KM tissue growth factor; immunomodulatory; cancer; leukaemia;  
KM nervous system disorders; arthritis; inflammation.

OS Homo sapiens.

PN WO200164835-A2.

PD 07-SEP-2001.

PF 26-FEB-2001; 2001WO-US04927.

PR 28-FEB-2000; 2000US-0515126.

PR 18-MAY-2000; 2000US-0577409.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT;

DR WPI: 2001-514838/56.

DR N-PSDB: AA189836.

PT Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders -

PS Claim 20: SEQ ID NO 23797; 1399pp + Sequence Listing; English.

CC The invention relates to human polynucleotides (AA179941-AA193841) and  
CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.

CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 64 AA;

Query Match 69.2%; Score 36; DB 22; Length 64;

Best Local Similarity 85.7%; Pred. No. 15;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SETINCH 7

DB 55 SETISCH 61

#### RESULT 15

ABP35700 ABP35700 standard; Protein; 863 AA.

AC ABP35700;

DT 24-JUL-2002 (first entry)

DE Fungal ZBC protein sequence #126.

KM Secondary metabolite; fungus; ZBC gene; zinc binuclear cluster protein;  
KM antibacterial; beta-lactam; anti-hypercholesterolaemic; lovastatin;  
KM mevastatin; immunosuppressant; cyclosporin A; ergot alkaloid; ergotamine;  
KM angiogenesis inhibitor; ovalicin; glucan synthase inhibitor; gliotoxin;  
KM fungal toxin; cell surface receptor; plant growth regulator; pigment;  
KM insecticide; antineoplastic.

OS Unidentified.

PN WO200224865-A2.

PD 28-MAR-2002.

PF 19-SEP-2001; 2001WO-US29288.

PR 19-SEP-2000; 2000US-233564P.

PA (MICR-) MICROBIA INC.

PI Holtzman D, Madden K, Maxon M, Sherman A;

DR WPI: 2002-352005/38.

DR N-PSDB: ABN79889.

PT New method for improving the production of a secondary metabolite e.g.  
PT antineoplastic agent, ergot alkaloid from a fungus involves modulation  
PT of the expression of at least one zinc binuclear cluster protein gene

PS Disclosure: SEQ ID 264; 49pp + sequence listing; English.

CC The invention relates to improving the production of a secondary  
CC metabolite by a fungus. This involves modulating the expression of at  
CC least one ZBC (zinc binuclear cluster protein) gene in a manner to  
CC improve the yield of the secondary metabolite. Methods of the invention  
CC may be used for improving the production of the secondary metabolite e.g.  
CC antibacterial (such as beta-lactam), an anti-hypercholesterolaemic (such  
CC as lovastatin or mevastatin), an immunosuppressant (such as cyclosporin A),  
CC an ergot alkaloid (such as ergotamine), an angiogenesis inhibitor (such  
CC as ovalicin), a glucan synthase inhibitor, gliotoxin family of compounds,  
CC a fungal toxin, a modulator of cell surface receptor signaling, a plant  
CC growth regulator, a pigment, an insecticide, or an antineoplastic  
CC compound. The method results in a decrease in fermentor run-time, a  
CC decrease in the size of the fermentor required for the production of  
CC equivalent amounts of the secondary metabolite, or a decrease in the  
CC biomass required for the production, which translates into decreased  
CC waste that must be handled in downstream processing. The sequences given  
CC in records ABP35575-ABP35722 represent ZBC proteins.

CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 863 AA;

Query Match 69.2%; Score 36; DB 23; Length 863;

Best Local Similarity 62.5%; Pred. No. 2.3e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 2 ETINCHFK 9  
:1:1111  
Db 470 KTLOCHFK 477

Search completed: July 11, 2003, 11:54:17  
Job time : 6.16424 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: July 11, 2003, 11:49:36 ; Search time 1.60915 Seconds

(without alignments)  
164.563 Million cell updates/sec

Title: US-10-091-442-1

Perfect score: 52

Sequence: 1 SETINCHFK 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Issued Patents-AA:\*  
1: /cgn2\_6/pdata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/pdata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/pdata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/pdata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/pdata/1/1aa/6CTUS.COMB.pep:\*  
6: /cgn2\_6/pdata/1/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	52	100.0	9	2	US-08-611-977-1
2	52	100.0	380	2	US-08-472-659-34
3	52	100.0	380	2	US-08-474-661-34
4	52	100.0	380	2	US-08-611-977-34
5	37	71.2	9	2	US-08-472-659-1
6	37	71.2	9	2	US-08-474-661-1
7	37	71.2	320	2	US-08-101-435-4
8	35	67.3	481	5	PCR-US91-02166-13
9	35	67.3	858	4	US-07-956-483-17
10	35	67.3	915	4	US-08-472-240A-8
11	34	65.4	185	2	US-08-463-911-3
12	34	65.4	195	3	US-08-955-937A-4
13	34	65.4	195	3	US-09-300-985-4
14	34	65.4	201	3	US-08-955-937A-2
15	34	65.4	201	3	US-09-300-985-2
16	34	65.4	236	4	US-09-140-804-6
17	34	65.4	804	4	US-09-067-091-2
18	34	65.4	1078	1	US-08-485-588-7
19	34	65.4	1078	1	US-08-484-565-7
20	34	65.4	1078	1	US-08-480-751-7
21	34	65.4	1078	2	US-08-943-986-7
22	34	65.4	1078	3	US-08-353-784-7
23	34	65.4	1078	3	US-08-484-719B-7
24	34	65.4	1078	4	US-08-484-159-7
25	34	65.4	1079	1	US-08-485-588-8
26	34	65.4	1079	1	US-08-484-565-8
27	34	65.4	1079	2	US-08-480-751-8

28	34	65.4	1079	2	US-08-943-986-8	Sequence 8, Appl1
29	34	65.4	1079	3	US-08-353-784-8	Sequence 8, Appl1
30	34	65.4	1079	3	US-08-484-719B-8	Sequence 8, Appl1
31	34	65.4	1079	4	US-08-484-159-8	Sequence 8, Appl1
32	34	65.4	1085	1	US-08-485-588-5	Sequence 5, Appl1
33	34	65.4	1085	1	US-08-484-565-5	Sequence 5, Appl1
34	34	65.4	1085	2	US-08-480-751-5	Sequence 5, Appl1
35	34	65.4	1085	2	US-08-943-986-5	Sequence 5, Appl1
36	34	65.4	1085	3	US-08-353-784-5	Sequence 5, Appl1
37	34	65.4	1085	3	US-08-484-719B-5	Sequence 5, Appl1
38	34	65.4	1085	4	US-08-484-159-5	Sequence 5, Appl1
39	34	65.4	1088	1	US-08-485-588-6	Sequence 6, Appl1
40	34	65.4	1088	1	US-08-484-565-6	Sequence 6, Appl1
41	34	65.4	1088	2	US-08-480-751-6	Sequence 6, Appl1
42	34	65.4	1088	2	US-08-943-986-6	Sequence 6, Appl1
43	34	65.4	1088	3	US-08-353-784-6	Sequence 6, Appl1
44	34	65.4	1088	3	US-08-484-719B-6	Sequence 6, Appl1
45	34	65.4	1088	4	US-08-484-159-6	Sequence 6, Appl1

#### ALIGNMENTS

RESULT 1  
US-08-611-977-1  
Sequence 1, Application US/08611977  
Patent No. 5972886  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSUROOKA, No. 5972886uo  
APPLICANT: NAKAZATO, Hitoshi  
APPLICANT: MURA, Kenju  
APPLICANT: ISHIDA, No. 5972886hito  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, No. 5972886omi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/611,977  
FILING DATE: 06-MAR-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-204  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-6021  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:

LENGTH: 9 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-611-977-1

Query Match 100.0%; Score 52; DB 2; Length 9;  
Best Local Similarity 100.0%; Pred. No. 2e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SETINCHFK 9  
1111111111  
DB 1 SETINCHFK 9

RESULT 2  
US-08-472-659-34  
Sequence 34, Application US/08472659  
Patent No. 5831030

GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSURUOKA, No. 5831030uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. 5831030uhlro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, No. 5831030mi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/472,659  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-248  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-472-659-34

Query Match 100.0%; Score 52; DB 2; Length 380;

Best Local Similarity 100.0%; Pred. No. 0.028;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SETINCHFK 9  
1111111111  
DB 188 SETINCHFK 196

RESULT 3  
US-08-474-661-34  
Sequence 34, Application US/08474661  
Patent No. 5874253

GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSURUOKA, No. 5874253uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. 5874253uhlro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, No. 5874253mi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/474,661  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: REA, TERESA STANER  
REGISTRATION NUMBER: 30,427  
REFERENCE/DOCKET NUMBER: 001560-204  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-6620  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-474-661-34

Query Match 100.0%; Score 52; DB 2; Length 380;  
Best Local Similarity 100.0%; Pred. No. 0.028;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SETINCHFK 9  
1111111111  
DB 188 SETINCHFK 196

RESULT 4

```

US-08-611-977-34
; Sequence 34, Application US/08611977
; Patent No. 5972886
; GENERAL INFORMATION:
; APPLICANT: TSUJIMOTO, Masafumi
; APPLICANT: IWASA, Fuyuki
; APPLICANT: TSUROUOKA, No. 5972886uo
; APPLICANT: NAKAZATO, Hiroshi
; APPLICANT: MIURA, Kenju
; APPLICANT: ISHIDA, No. 5972886uhlro
; APPLICANT: KURIHARA, Tatsuya
; APPLICANT: YAMAUCHI, Kozo
; TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/611.977
; FILING DATE: 06-MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/091,028
; FILING DATE: 14-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 4-212305
; FILING DATE: 17-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6-067339
; FILING DATE: 04-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 001560-204
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 380 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-611-977-34

Query Match          100.0%  Score 52;  DB 2;  Length 380;
Best Local Similarity 100.0%; Pred. No. 0.028;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 SETINCHFK 9
DB      188 SETINCHFK 196

RESULT 5
US-08-472-659-1
; Sequence 1, Application US/08472659
; Patent No. 5831030
; GENERAL INFORMATION:
; APPLICANT: TSUJIMOTO, Masafumi
; APPLICANT: IWASA, Fuyuki
; APPLICANT: TSUROUOKA, No. 5831030uo
; APPLICANT: NAKAZATO, Hiroshi

```

```

; APPLICANT: MIURA, Kenju
; APPLICANT: ISHIDA, No. 5831030uhlro
; APPLICANT: KURIHARA, Tatsuya
; APPLICANT: YAMAUCHI, Kozo
; APPLICANT: YAMAUCHI, No. 5831030oml
; TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,659
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 4-212305
; FILING DATE: 17-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 5-067339
; FILING DATE: 04-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/091,028
; FILING DATE: 14-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 001560-248
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-472-659-1

Query Match          71.2%  Score 37;  DB 2;  Length 9;
Best Local Similarity 87.5%; Pred. No. 2e+05;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      2 ETINCHFK 9
DB      2 ETINCHFK 9

RESULT 6
US-08-474-661-1
; Sequence 1, Application US/08474661
; Patent No. 5874253
; GENERAL INFORMATION:
; APPLICANT: TSUJIMOTO, Masafumi
; APPLICANT: IWASA, Fuyuki
; APPLICANT: TSUROUOKA, No. 5874253uo
; APPLICANT: NAKAZATO, Hiroshi
; APPLICANT: MIURA, Kenju
; APPLICANT: ISHIDA, No. 5874253uhlro
; APPLICANT: KURIHARA, Tatsuya
; APPLICANT: YAMAUCHI, Kozo
; APPLICANT: YAMAUCHI, No. 5874253oml
; TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR
; NUMBER OF SEQUENCES: 34

```

;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Burns, Doane, Swecker & Mathis  
;; STREET: George Mason Bldg., Washington & Prince Sts.  
;; CITY: Alexandria  
;; STATE: Virginia  
;; COUNTRY: United States  
;; ZIP: 22313-1404  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.25  
;;  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/474,661  
;; FILING DATE: 07-JUN-1995  
;; CLASSIFICATION: 435  
;;  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/091,028  
;; FILING DATE: 14-JUL-1993  
;; APPLICATION NUMBER: JP 4-212305  
;; FILING DATE: 17-JUL-1992  
;;  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: JP 5-067339  
;; FILING DATE: 04-MAR-1993  
;;  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: REA, TERESA STANER  
;; REGISTRATION NUMBER: 30,427  
;; REFERENCE/DOCKET NUMBER: 001560-204  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (703) 836-6620  
;; TELEFAX: (703) 836-6620  
;;  
;; INFORMATION FOR SEQ ID NO: 1:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 9 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;;  
;; MOLECULE TYPE: peptide  
;;  
;; US-08-474-661-1

Query Match 71.2%; Score 37; DB 2; Length 9;  
Best Local Similarity 87.5%; Pred. No. 2e+05;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 ETINCHFK 9  
      |||||  
Db 2 ETINCHFK 9

RESULT 7  
US-08-101-435-4  
Sequence 4, Application US/08101435  
Patent No. 5441883  
GENERAL INFORMATION:  
APPLICANT: Civeill, Olivier  
APPLICANT: Zhou, Qun-Yong  
TITLE OF INVENTION: A No. 5441883el Adenosine Receptor and Uses  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Allegretti & Witcoff, Ltd.  
STREET: 10 South Wacker Drive, Suite 3000  
CITY: Chicago  
STATE: IL  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/101,435  
FILING DATE:

;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US/07/847,563  
;; FILING DATE:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: McDonnell, John J  
;; REGISTRATION NUMBER: 26,949  
;; REFERENCE/DOCKET NUMBER: 91,708  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 312-715-1000  
;; TELEFAX: 312-715-1234  
;; TELEX: 910-221-5317  
;;  
;; INFORMATION FOR SEQ ID NO: 4:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 320 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;;  
;; MOLECULE TYPE: protein  
;;  
;; US-08-101-435-4

Query Match 71.2%; Score 37; DB 1; Length 320;  
Best Local Similarity 55.6%; Pred. No. 14;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 SETINCHFK 9  
      |::|||:  
Db 163 SSTLSCHFK 171

RESULT 8  
PCT-US91-02166-13  
Sequence 13, Application PC/TUS9102166  
GENERAL INFORMATION:  
APPLICANT: Genentech, Inc.  
TITLE OF INVENTION: HIV Envelope Polypeptides  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 KB floppy disk  
COMPUTER: IBM pc compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: pacin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/02166  
FILING DATE: 19910401  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: U.S.S.N. 07/504,772  
FILING DATE: 03-APRIL-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Adler, Carolyn R.  
REGISTRATION NUMBER: 32,324  
REFERENCE/DOCKET NUMBER: 639  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/266-2614  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 481 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
PCT-US91-02166-13

Query Match 67.3%; Score 35; DB 5; Length 481;  
Best Local Similarity 85.7%; Pred. No. 52;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;



OY 2 ETINCH 8  
11111  
Db 140 ETINCOF 146

RESULT 9  
US-07-956-483-17  
Sequence 17, Application US/07956483  
Patent No. 6261799  
GENERAL INFORMATION:  
APPLICANT: KIENY, Marie-Paule  
TITLE OF INVENTION: NOVEL HYBRID, SOLUBLE AND UNCLEAVED  
TITLE OF INVENTION: 9p160 VARIANT  
NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/956,483  
FILING DATE: 31-DEC-1992  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: WO 92/19742  
FILING DATE: 12-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 91 05392  
FILING DATE: 02-MAY-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Crane-Feury, Sharon E  
REGISTRATION NUMBER: 36,113  
REFERENCE/DOCKET NUMBER: 017753-005  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ. ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 858 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-07-956-483-17  
Query Match 67.3%; Score 35; DB 4; Length 858;  
Best Local Similarity 85.7%; Pred. No. 95;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 ETINCH 8  
11111  
Db 159 ETINCOF 165

RESULT 10  
US-08-472-240A-8  
Sequence 8, Application US/08472240A  
Patent No. 6284248  
GENERAL INFORMATION:  
APPLICANT: KIENY, Marie-Paule  
TITLE OF INVENTION: NOVEL HYBRID, SOLUBLE AND UNCLEAVED  
TITLE OF INVENTION: 9p160 VARIANT  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS

STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/472,240A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/956,483  
FILING DATE: 31-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Teskin, Robin L.  
REGISTRATION NUMBER: 35,030  
REFERENCE/DOCKET NUMBER: 017753-055  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ. ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 915 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..883  
US-08-472-240A-8

Query Match 67.3%; Score 35; DB 4; Length 915;  
Best Local Similarity 85.7%; Pred. No. 16+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 ETINCH 8  
11111  
Db 182 ETINCOF 188

RESULT 11  
US-08-463-911-3  
Sequence 3, Application US/08463911  
Patent No. 5869330  
GENERAL INFORMATION:  
APPLICANT: Scherer, Philipp E.  
TITLE OF INVENTION: A NOVEL SERUM PROTEIN PRODUCED  
TITLE OF INVENTION: EXCLUSIVELY IN ADIPOCYTES  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/463,911  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia

REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: WHI95-05  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 861-6240  
TELEFAX: (617) 861-9540  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 185 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-463-911-3

Query Match 65.4%; Score 34; DB 2; Length 185;  
Best Local Similarity 71.4%; Pred. No. 29;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 TINCHK 9  
DB 51 TVNCHK 57

RESULT 12  
US-08-955-937A-4  
Sequence 4, Application US/08955937A  
Patent No. 6020161  
GENERAL INFORMATION:  
APPLICANT: WU, SHUTIAN  
APPLICANT: SWEET, RAYMOND  
APPLICANT: TRUNEH, ALEMSSEG  
TITLE OF INVENTION: PIGR-1, A MEMBER OF IMMUNOGLOBULIN  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: RATNER & PRESTIA  
STREET: P.O. BOX 980  
CITY: VALLEY FORGE  
STATE: PA  
COUNTRY: USA  
ZIP: 19482  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/955,937A  
FILING DATE: 17-OCT-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/056,152  
FILING DATE: 19-AUG-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: PRESTIA, PAUL F  
REGISTRATION NUMBER: 23,031  
REFERENCE/DOCKET NUMBER: GH-70228  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-407-0700  
TELEFAX: 610-407-0701  
TELEX: 846169  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 195 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-955-937A-4

Query Match 65.4%; Score 34; DB 3; Length 195;  
Best Local Similarity 55.6%; Pred. No. 31;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 SETINCHK 9  
DB 31 SLTVQCHK 39

RESULT 13  
US-09-300-985-4  
Sequence 4, Application US/09300985A  
Patent No. 6232441  
GENERAL INFORMATION:  
APPLICANT: WU, SHUTIAN  
APPLICANT: SWEET, RAYMOND W.  
APPLICANT: TRUNEH, ALEMSSEG  
APPLICANT: HURLE, MARK ROBERT  
TITLE OF INVENTION: PIGR-1, A MEMBER OF IMMUNOGLOBULIN GENE  
FILE REFERENCE: GH-70228-1  
CURRENT APPLICATION NUMBER: US/09/300,985A  
CURRENT FILING DATE: 1999-04-28  
EARLIER APPLICATION NUMBER: US 08/955,937  
EARLIER FILING DATE: 1997-10-22  
EARLIER APPLICATION NUMBER: US 60/056,152  
EARLIER FILING DATE: 1997-08-19  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: FASTSEQ for Windows Version 3.0  
SEQ ID NO 4  
LENGTH: 195  
TYPE: PRT  
ORGANISM: HOMO SAPIENS  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (122)(170)(184)  
US-09-300-985-4

Query Match 65.4%; Score 34; DB 4; Length 195;  
Best Local Similarity 55.6%; Pred. No. 31;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 SETINCHK 9  
DB 31 SLTVQCHK 39

RESULT 14  
US-08-955-937A-2  
Sequence 2, Application US/08955937A  
Patent No. 6020161  
GENERAL INFORMATION:  
APPLICANT: WU, SHUTIAN  
APPLICANT: SWEET, RAYMOND  
APPLICANT: TRUNEH, ALEMSSEG  
TITLE OF INVENTION: PIGR-1, A MEMBER OF IMMUNOGLOBULIN  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: RATNER & PRESTIA  
STREET: P.O. BOX 980  
CITY: VALLEY FORGE  
STATE: PA  
COUNTRY: USA  
ZIP: 19482  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/955,937A  
FILING DATE: 17-OCT-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/056,152

FILING DATE: 19-AUG-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: PRESTIA, PAUL F  
REGISTRATION NUMBER: 23,031  
REFERENCE/DOCKET NUMBER: GH-70228  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-407-0700  
TELEFAX: 610-407-0701  
TELEX: 846169  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 201 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-955-937A-2

Query Match 65.4%; Score 34; DB 3; Length 201;  
Best Local Similarity 55.6%; Pred. No. 32;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 SETINCHFK 9  
| | : | : |  
DB 31 SLTVQCHYK 39

RESULT 15  
US-09-300-985-2  
Sequence 2, Application US/09300985A  
Patent No. 6232441  
GENERAL INFORMATION:  
APPLICANT: WU, SHUTIAN  
APPLICANT: SWEET, RAYMOND W.  
APPLICANT: TRUNER, ALEKSEGED  
APPLICANT: HURLE, MARK ROBERT  
TITLE OF INVENTION: PIGR-1, A MEMBER OF IMMUNOGLOBULIN GENE  
FILE REFERENCE: GH-70228-1  
CURRENT APPLICATION NUMBER: US/09/300,985A  
CURRENT FILING DATE: 1999-04-28  
EARLIER APPLICATION NUMBER: US 08/955,937  
EARLIER FILING DATE: 1997-10-22  
EARLIER APPLICATION NUMBER: US 60/056,152  
EARLIER FILING DATE: 1997-08-19  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: fastseq for windows version 3.0  
SEQ ID NO 2  
LENGTH: 201  
TYPE: PRT  
ORGANISM: HOMO SAPIENS  
US-09-300-985-2

Query Match 65.4%; Score 34; DB 4; Length 201;  
Best Local Similarity 55.6%; Pred. No. 32;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 SETINCHFK 9  
| | : | : |  
DB 31 SLTVQCHYK 39

Search completed: July 11, 2003, 12:02:04  
Job time : 2.60915 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: July 11, 2003, 11:59:17 ; Search time 2.63825 Seconds  
(without alignments)  
397.148 Million cell updates/sec

Title: US-10-091-442-1  
Perfect score: 52  
Sequence: 1 SERINCHFK 9

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 445758 seqs, 116419773 residues

Total number of hits satisfying chosen parameters: 445758

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

Published\_Applications\_AA:\*  
1: /cgn2\_6/ptodata/1/pubppaa/US07\_NEW\_PUB.pep:\*  
2: /cgn2\_6/ptodata/1/pubppaa/PCITUS\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubppaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubppaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubppaa/PCITUS\_PUBCOMB.pep:\*  
6: /cgn2\_6/ptodata/1/pubppaa/US08\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubppaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubppaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubppaa/US09\_NEW\_PUB.pep:\*  
10: /cgn2\_6/ptodata/1/pubppaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubppaa/US10\_NEW\_PUB.pep:\*  
12: /cgn2\_6/ptodata/1/pubppaa/US10\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/1/pubppaa/US60\_NEW\_PUB.pep:\*  
14: /cgn2\_6/ptodata/1/pubppaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	52	100.0	9	US-10-091-442-1	Sequence 1, Appli
2	52	100.0	9	US-09-140-719-1	Sequence 1, Appli
3	52	100.0	380	US-10-091-442-34	Sequence 34, Appli
4	52	100.0	380	US-09-140-719-34	Sequence 34, Appli
5	35	67.3	90	US-09-764-881-113	Sequence 113, App
6	35	67.3	90	US-09-764-853-616	Sequence 616, App
7	34	65.4	201	US-10-174-590-608	Sequence 608, App
8	34	65.4	201	US-10-176-758-608	Sequence 608, App
9	34	65.4	201	US-10-175-737-608	Sequence 608, App
10	34	65.4	201	US-10-173-706-608	Sequence 608, App
11	34	65.4	201	US-10-175-738-608	Sequence 608, App
12	34	65.4	201	US-10-175-752-608	Sequence 608, App
13	34	65.4	201	US-10-176-482-608	Sequence 608, App
14	34	65.4	201	US-10-176-757-608	Sequence 608, App
15	34	65.4	201	US-10-176-913-608	Sequence 608, App
16	34	65.4	201	US-10-180-552-608	Sequence 608, App
17	34	65.4	201	US-10-180-557-608	Sequence 608, App
18	34	65.4	201	US-10-173-700-608	Sequence 608, App
19	34	65.4	201	US-10-174-572-608	Sequence 608, App

20	34	65.4	201	US-10-174-579-608	Sequence 608, App
21	34	65.4	201	US-10-174-582-608	Sequence 608, App
22	34	65.4	201	US-10-174-588-608	Sequence 608, App
23	34	65.4	201	US-10-175-739-608	Sequence 608, App
24	34	65.4	201	US-10-175-740-608	Sequence 608, App
25	34	65.4	201	US-10-175-743-608	Sequence 608, App
26	34	65.4	201	US-10-176-488-608	Sequence 608, App
27	34	65.4	201	US-10-176-492-608	Sequence 608, App
28	34	65.4	201	US-10-176-747-608	Sequence 608, App
29	34	65.4	201	US-10-176-750-608	Sequence 608, App
30	34	65.4	201	US-10-176-985-608	Sequence 608, App
31	34	65.4	201	US-10-176-987-608	Sequence 608, App
32	34	65.4	201	US-10-176-991-608	Sequence 608, App
33	34	65.4	201	US-10-176-992-608	Sequence 608, App
34	34	65.4	201	US-10-176-993-608	Sequence 608, App
35	34	65.4	201	US-10-184-658-608	Sequence 608, App
36	34	65.4	201	US-10-173-635-608	Sequence 608, App
37	34	65.4	201	US-10-173-697-608	Sequence 608, App
38	34	65.4	201	US-10-173-705-608	Sequence 608, App
39	34	65.4	201	US-10-174-576-608	Sequence 608, App
40	34	65.4	201	US-10-174-585-608	Sequence 608, App
41	34	65.4	201	US-10-174-586-608	Sequence 608, App
42	34	65.4	201	US-10-175-747-608	Sequence 608, App
43	34	65.4	201	US-10-176-481-608	Sequence 608, App
44	34	65.4	201	US-10-176-485-608	Sequence 608, App
45	34	65.4	201	US-10-176-487-608	Sequence 608, App

## ALIGNMENTS

RESULT 1  
US-10-091-442-1  
Sequence 1, Application US/10091442  
Patent No. US20020164711A1  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
IWASA, Fuyuki  
TSURUOKA, No. US20020164711A1uo  
NAKAZATO, Hiroshi  
MURA, Kenju  
ISHIDA, No. US20020164711A1uhfro  
KURIHARA, Tatsuya  
YAMACHI, Koza  
YAMAGUCHI, No. US20020164711A1omf  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/091,442  
FILING DATE: 07-Mar-2002  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-Aug-1998  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-Jun-1995  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-Jul-1993  
APPLICATION NUMBER: JP 4-213305  
FILING DATE: 17-Jul-1992  
APPLICATION NUMBER: JP 6-067339

FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-10-091-442-1

Query Match 100.0%; Score 52; DB 9; Length 9;  
Best Local Similarity 100.0%; Pred. No. 4e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SETINCHK 9  
Db 1 SETINCHK 9

RESULT 2  
US-09-140-719-1  
Sequence 1, Application US/09140719  
Patent No. US20010026931A1  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSUROOKA, No. US20010026931A1no  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MURA, Kenju  
APPLICANT: ISHIDA, No. US20010026931A1uhiro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAGUCHI, Kozo  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.

REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-140-719-1

Query Match 100.0%; Score 52; DB 10; Length 9;  
Best Local Similarity 100.0%; Pred. No. 4e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SETINCHK 9  
Db 1 SETINCHK 9

RESULT 3  
US-10-091-442-34  
Sequence 34, Application US/10091442  
Patent No. US20020164711A1  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSUROOKA, No. US20020164711A1no  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MURA, Kenju  
APPLICANT: ISHIDA, No. US20020164711A1uhiro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAGUCHI, Kozo  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/091,442  
FILING DATE: 07-MAR-2002  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021

INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 34:  
US-10-091-442-34

Query Match 100.0%; Score 52; DB 9; Length 380;  
Best Local Similarity 100.0%; Pred. No. 0.09;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SETINCHK 9  
DB 188 SETINCHK 196

RESULT 4  
US-09-140-719-34  
Sequence 34, Application US/09140719  
Patent No. US20010026931A1  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASHI, Fuyuki  
APPLICANT: TSURUOKA, No. US20010026931A1no  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. US20010026931A1uhlro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAUCHI, Kozo  
APPLICANT: YAMAGUCHI, No. US20010026931A1oml  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid

TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-140-719-34

Query Match 100.0%; Score 52; DB 10; Length 380;  
Best Local Similarity 100.0%; Pred. No. 0.09;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SETINCHK 9  
DB 188 SETINCHK 196

RESULT 5  
US-09-764-881-113  
Sequence 113, Application US/09764881  
Publication No. US20030125246A9  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PT207  
CURRENT APPLICATION NUMBER: US/09/764,881  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - refer to PALM or file wrapper  
NUMBER OF SEQ ID NOS: 192  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 113  
LENGTH: 90  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (20)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-881-113

Query Match 67.3%; Score 35; DB 9; Length 90;  
Best Local Similarity 77.8%; Pred. No. 27;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 SETINCHK 9  
DB 69 SETINCHK 77

RESULT 6  
US-09-764-853-616  
Sequence 616, Application US/09764853  
Patent No. US20020090672A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PJ206  
CURRENT APPLICATION NUMBER: US/09/764,853  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 939  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 616  
LENGTH: 90  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (20)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-853-616

Query Match 67.3%; Score 35; DB 10; Length 90;  
Best Local Similarity 77.8%; Pred. No. 27;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 SETINCHK 9

Db 69 SETINCVLK 77

## RESULT 7

US-10-174-590-608  
; Sequence 608, Application US/10174590  
; Publication No. US20030008352A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3430R1C42  
; CURRENT APPLICATION NUMBER: US/10/174,590  
; CURRENT FILING DATE: 2002-06-18  
; Prior Application removed - See file Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 612  
; SEQ ID NO 608  
; LENGTH: 201  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-174-590-608

Query Match 65.4%; Score 34; DB 9; Length 201;  
Best Local Similarity 55.6%; Pred. No. 90;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 SETINCHK 9  
Db 31 SLTVQCHXK 39

## RESULT 8

US-10-176-758-608  
; Sequence 608, Application US/10176758  
; Publication No. US20030008353A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3430R1C104  
; CURRENT APPLICATION NUMBER: US/10/176,758  
; CURRENT FILING DATE: 2002-06-21  
; Prior Application removed - See file Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 612  
; SEQ ID NO 608  
; LENGTH: 201  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-176-758-608

Query Match 65.4%; Score 34; DB 9; Length 201;  
Best Local Similarity 55.6%; Pred. No. 90;

Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
OY 1 SETINCHK 9  
Db 31 SLTVQCHXK 39

## RESULT 9

US-10-175-737-608  
; Sequence 608, Application US/10175737  
; Publication No. US20030013153A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3430R1C50  
; CURRENT APPLICATION NUMBER: US/10/175,737  
; CURRENT FILING DATE: 2002-06-19  
; Prior Application removed - See file Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 612  
; SEQ ID NO 608  
; LENGTH: 201  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-175-737-608

Query Match 65.4%; Score 34; DB 9; Length 201;  
Best Local Similarity 55.6%; Pred. No. 90;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 SETINCHK 9  
Db 31 SLTVQCHXK 39

## RESULT 10

US-10-173-706-608  
; Sequence 608, Application US/10173706  
; Publication No. US2003002293A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3430R1C7  
; CURRENT APPLICATION NUMBER: US/10/173,706  
; CURRENT FILING DATE: 2002-06-17  
; Prior Application removed - See file Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 612  
; SEQ ID NO 608  
; LENGTH: 201  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-173-706-608



```
Query Match          65.4%; Score 34; DB 9; Length 201;
Best Local Similarity 55.6%; Pred. No. 90;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 SETINCHK 9
   1 1 1 1 1 1
Db 31 SLTVOCHK 39

RESULT 11
US-10-175-738-608
; Sequence 608, Application US/10175738
; Publication No. US20030022294A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C45
; CURRENT APPLICATION NUMBER: US/10/175,738
; CURRENT FILING DATE: 2002-06-19
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 608
; LENGTH: 201
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-738-608

Query Match          65.4%; Score 34; DB 9; Length 201;
Best Local Similarity 55.6%; Pred. No. 90;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 SETINCHK 9
   1 1 1 1 1 1
Db 31 SLTVOCHK 39

RESULT 12
US-10-175-752-608
; Sequence 608, Application US/10175752
; Publication No. US20030022295A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C60
; CURRENT APPLICATION NUMBER: US/10/175,752
; CURRENT FILING DATE: 2002-06-19
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 608
; LENGTH: 201
```

```
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-752-608

Query Match          65.4%; Score 34; DB 9; Length 201;
Best Local Similarity 55.6%; Pred. No. 90;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 SETINCHK 9
   1 1 1 1 1 1
Db 31 SLTVOCHK 39

RESULT 13
US-10-176-482-608
; Sequence 608, Application US/10176482
; Publication No. US20030022296A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C70
; CURRENT APPLICATION NUMBER: US/10/176,482
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 608
; LENGTH: 201
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-482-608

Query Match          65.4%; Score 34; DB 9; Length 201;
Best Local Similarity 55.6%; Pred. No. 90;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 SETINCHK 9
   1 1 1 1 1 1
Db 31 SLTVOCHK 39

RESULT 14
US-10-176-757-608
; Sequence 608, Application US/10176757
; Publication No. US20030022297A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C86
; CURRENT APPLICATION NUMBER: US/10/176,757
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
```

; NUMBER OF SEQ ID NOS: 612  
; SEQ ID NO 608  
; LENGTH: 201  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-176-757-608

Query Match 65.4%; Score 34; DB 9; Length 201;  
Best Local Similarity 55.6%; Pred. No. 90;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 SETINCHK 9  
| : | : | : |  
Db 31 SLTVQCHYK 39

RESULT 15  
US-10-176-913-608  
; Sequence 608, Application US/10176913  
; Publication No. US20030022298A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; TITLE OF INVENTION: ACIDS ENCODING THE SAME  
; FILE REFERENCE: P430R1C66  
; CURRENT APPLICATION NUMBER: US/10/176,913  
; CURRENT FILING DATE: 2002-06-20  
; Prior Application removed - See file Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 612  
; SEQ ID NO 608  
; LENGTH: 201  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-176-913-608

Query Match 65.4%; Score 34; DB 9; Length 201;  
Best Local Similarity 55.6%; Pred. No. 90;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 SETINCHK 9  
| : | : | : |  
Db 31 SLTVQCHYK 39

Search completed: July 11, 2003, 12:37:31  
Job time : 3.63825 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:45:06 ; Search time 1.81497 Seconds  
(without alignments)  
476.708 Million cell updates/sec

Title: US-10-091-442-1

Perfect score: 52

Sequence: 1 SETINCHFR 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : PIR73:\*

1: PIR1:\*

2: PIR2:\*

3: PIR3:\*

4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	71.2	320	2	S17177
2	37	71.2	320	2	A46152
3	37	71.2	337	2	S68678
4	37	71.2	1030	2	S57380
5	36	69.2	863	2	S38140
6	35	67.3	118	2	B64028
7	35	67.3	203	2	I49054
8	35	67.3	267	2	I49053
9	35	67.3	267	2	I55686
10	35	67.3	280	2	I49052
11	35	67.3	280	2	D84968
12	35	67.3	426	2	F95058
13	35	67.3	522	2	H97927
14	35	67.3	858	1	VCLJG2
15	35	67.3	1173	2	T25539
16	35	67.3	2095	2	S29529
17	34	65.4	82	2	T10149
18	34	65.4	133	2	I46632
19	34	65.4	136	2	I46635
20	34	65.4	138	2	I46634
21	34	65.4	145	2	I46633
22	34	65.4	215	2	C48150
23	34	65.4	231	1	HLCHBL
24	34	65.4	256	2	T28853
25	34	65.4	389	2	F81211
26	34	65.4	389	2	A81787
27	34	65.4	779	2	A86862
28	34	65.4	798	2	C95202
29	34	65.4	798	2	C98069

30	34	65.4	1078	2	A56715	calcium receptor (
31	34	65.4	1079	2	I59362	calcium/polyvalent
32	34	65.4	1085	2	S40476	Ca(2+)-sensing rec
33	34	65.4	1088	2	B56715	calcium receptor (
34	33	63.5	59	2	A62023	hypothetical prote
35	33	63.5	129	2	B48909	adenosine receptor
36	33	63.5	131	2	T29743	hypothetical prote
37	33	63.5	169	2	T33949	hypothetical prote
38	33	63.5	226	2	T39298	hypothetical prote
39	33	63.5	288	2	H75069	cytochrome-c3 hyd
40	33	63.5	298	2	T33220	hypothetical prote
41	33	63.5	313	2	A48903	beta-lactamase - P
42	33	63.5	320	2	E71696	rare lipoprotein A
43	33	63.5	348	2	D89102	protein F25E5.2 (l
44	33	63.5	347	2	T20472	hypothetical prote
45	33	63.5	367	2	T18839	hypothetical prote

#### ALIGNMENTS

##### RESULT 1

S17177

probable G protein-coupled receptor - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 04-Dec-1992 #sequence-revision 04-Dec-1992 #text\_change 29-Sep-1999

C:Accession: S17177

R:Meyerhof, W.; Mueller-Brechlin, R.; Richter, D.

FEBS Lett. 284, 155-160, 1991

A>Title: Molecular cloning of a novel putative G-protein coupled receptor expressed d

A:Reference number: S17177; MUID:91285122; PMID:1647979

A:Accession: S17177

A:Molecule type: mRNA

A:Residues: 1-320 <MEY>

A:Cross-references: EMBL:X59249; NID:956307; PIDN:CAA41937.1; PID:956308

C:Superfamily: adenosine receptor A1

C:Keywords: G protein-coupled receptor; glycoprotein; lipoprotein; thiolester bond; t

F:17-79/Domain: transmembrane #status predicted <TM1>

F:51-74/Domain: transmembrane #status predicted <TM2>

F:84-108/Domain: transmembrane #status predicted <TM3>

F:129-150/Domain: transmembrane #status predicted <TM4>

F:180-200/Domain: transmembrane #status predicted <TM5>

F:234-257/Domain: transmembrane #status predicted <TM6>

F:264-286/Domain: transmembrane #status predicted <TM7>

F:4,5,42,162,213/Binding site: carbonylate (asn) (covalent) #status predicted

F:302/Binding site: palmitate (cys) (covalent) #status predicted

Query Match 71.2% Score 37; DB 2; Length 320;

Best local similarity 55.6% Pred. No. 25; Mismatches 1; Indels 0; Gaps 0;

Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 SETINCHFR 9

Db 163 SSTLSCHFR 171

##### RESULT 2

A46152

A3 adenosine receptor - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 21-Sep-1993 #sequence-revision 18-Nov-1994 #text\_change 29-Sep-1999

C:Accession: A46152

R:Zhong, Q.Y.; Li, C.; Olah, M.E.; Johnson, R.A.; Stiles, G.L.; Clivell, O.

Proc. Natl. Acad. Sci. U.S.A. 89, 7432-7436, 1992

A>Title: Molecular cloning and characterization of an adenosine receptor: the A3 aden

A:Reference number: A46152; MUID:92366475; PMID:1323836

A:Accession: A46152

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-320 <ZHO>

A:Cross-references: GB:M94152; NID:9202724; PIDN:AAA40680.1; PID:9460332

A:Experimental source: brain

A>Note: sequence inconsistent with the nucleotide translation

A:Note: sequence extracted from NCBI backbone (NCBIN:110661, NCBI:110662)  
C:Superfamily: adenosine receptor A1  
C:Keywords: G protein-coupled receptor; transmembrane protein

Query Match 71.2%; Score 37; DB 2; Length 320;  
Best Local Similarity 55.6%; Pred. No. 25;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 SETINCHFR 9  
Db 163 STLSCHFR 171

## RESULT 3

adenosine receptor A31 - rat  
C:Species: Rattus sp. (rat)  
C:Date: 15-Feb-1997 #sequence\_revision 13-Mar-1997 #text\_change 20-Jun-2000  
C:Accession: S68678  
R:SaJJadi, F.G.; Boyle, D.L.; Domingo, R.C.; Firestein, G.S.  
FEBS Lett. 382, 125-129, 1996  
A:Title: cDNA cloning and characterization of A31, an alternatively spliced rat A3 adenc  
A:Reference number: S68678; MUID:96196578; PMID:8612733  
A:Accession: S68678  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-337 <SAAV>  
A:Cross-references: EMBL:X93219; NID:g1297000; PIDN:CA63702.1; PID:g1297001  
C:Superfamily: adenosine receptor A1  
C:Keywords: alternative splicing

Query Match 71.2%; Score 37; DB 2; Length 337;  
Best Local Similarity 55.6%; Pred. No. 26;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 SETINCHFR 9  
Db 180 STLSCHFR 188

## RESULT 4

probable membrane protein YOL089c - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein 00938  
C:Species: Saccharomyces cerevisiae  
C:Date: 28-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 19-Apr-2002  
C:Accession: S57380; S66783; S50416  
R:Zumstein, E.; Pearson, B.M.; Kalogeropoulos, A.; Schweizer, M.  
Yeast 11, 975-986, 1995  
A:Title: A 29,425 kb segment on the left arm of yeast chromosome XV contains more than t  
A:Reference number: S57374; MUID:96021609; PMID:8533473  
A:Accession: S57380  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-1030 <ZUM>  
A:Cross-references: EMBL:X83121; NID:9600461; PIDN:CA58190.1; PID:9600469  
R:Zumstein, E.; Pearson, B.M.; Kalogeropoulos, A.; Schweizer, M.  
submitted to the Protein Sequence Database, July 1996.  
A:Reference number: S66775  
A:Accession: S66783  
A:Molecule type: DNA  
A:Residues: 1-1030 <ZUM>  
A:Cross-references: EMBL:Z74831; NID:g1419931; PIDN:CA99101.1; PID:e251886; PID:g141993  
A:Experimental source: strain S288C  
C:Genetics:

A:Gene: SCD:HAL9  
A:Cross-references: SGD:S0005449  
A:Map position: 15L  
C:Superfamily: unassigned GAL4-type zinc cluster proteins; GAL4 zinc binuclear cluster  
C:Keywords: transmembrane protein  
F:131-171/Domain: GAL4 zinc binuclear cluster homology <GAL4>  
F:762-778/Domain: transmembrane #status predicted <TM>

Query Match 71.2%; Score 37; DB 2; Length 1030;  
Best Local Similarity 55.6%; Pred. No. 69;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 SETINCHFR 9  
Db 689 SRVINCCHFR 697

## RESULT 5

probable finger protein YKR064w - yeast (Saccharomyces cerevisiae)  
C:Species: Saccharomyces cerevisiae  
C:Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 19-Apr-2002  
C:Accession: S38140  
R:van Vliet-Reedijk, J.C.; Planta, R.J.  
submitted to the Protein Sequence Database, March 1994  
A:Reference number: S38130  
A:Accession: S38140  
A:Molecule type: DNA  
A:Residues: 1-863 <VAN>  
A:Cross-references: EMBL:Z28289; NID:g486530; PIDN:CA82143.1; PID:g486531; MIPS:YKR0  
A:Experimental source: strain S288C  
C:Genetics:  
A:Cross-references: SGD:S0001772  
A:Map position: 11R  
C:Superfamily: unassigned GAL4-type zinc cluster proteins; GAL4 zinc binuclear cluste  
C:Keywords: DNA binding; nucleic acid transcription regulation; zinc finger  
F:14-52/Domain: GAL4 zinc binuclear cluster homology <GAL4>  
F:19-47/Region: zinc finger

Query Match 69.2%; Score 36; DB 2; Length 863;  
Best Local Similarity 62.5%; Pred. No. 90;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 ETINCHFR 9  
Db 470 KTLQCHFR 477

## RESULT 6

hypothetical protein H1406 - Haemophilus influenzae (strain Rd KW20)  
C:Species: Haemophilus influenzae  
C:Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 08-Oct-1999  
C:Accession: B64028  
R:Reischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage  
; Goodyne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodok, A.; Kelley, J.M.; Weidman  
; D.M.; Brindon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.  
Science 269, 496-512, 1995  
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vente  
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A:Reference number: A64000; MUID:95350630; PMID:7542800  
A:Accession: B64028  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-118 <TRGR>  
A:Cross-references: GB:U32820; GB:LA2023; NID:g1574231; PIDN:AC23056.1; PID:g1574244  
C:Superfamily: Haemophilus influenzae hypothetical protein H1406

Query Match 67.3%; Score 35; DB 2; Length 118;  
Best Local Similarity 66.7%; Pred. No. 25;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 SETINCHFR 9  
Db 35 SETVSCHFR 43

## RESULT 7

Ly-49G.2 antigen - mouse  
C:Species: Mus musculus (house mouse)

```
C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 17-Mar-2000
C:Accession: I49054
R:Smith, H.R.; Karhofer, F.M.; Yokoyama, W.M.
J. Immunol. 153, 1068-1079, 1994
A:Title: Ly-49 multigene family expressed by IL-2-activated NK cells.
A:Reference number: I49049; MUID:94300068; PMID:8027540
A:Accession: I49054
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-203 <RES>
A:Cross-references: EMBL:U10095; NID:9533497; PIDN:AAA50223.1; PID:9533498
C:Superfamily: natural killer cell receptor P1; C-type lectin homology

Query Match          67.3%; Score 35; DB 2; Length 203;
Best Local Similarity 83.3%; Pred. NO. 39;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY      2 ETINCH 7
DB      80 ETLNCH 85

RESULT 8
I49053
Ly-49G.2 antigen - mouse
N:Alternate names: Ly-49E
C:Species: Mus musculus (house mouse)
C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 01-Dec-2000
C:Accession: I49053; I49115
R:Smith, H.R.; Karhofer, F.M.; Yokoyama, W.M.
J. Immunol. 153, 1068-1079, 1994
A:Title: Ly-49 multigene family expressed by IL-2-activated NK cells.
A:Reference number: I49049; MUID:94300068; PMID:8027540
A:Accession: I49053
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-267 <RES>
A:Cross-references: EMBL:U10094; NID:9533495; PIDN:AAA50222.1; PID:9533496
R:Brennan, J.; Mager, D.; Jefferys, W.; Takel, F.
J. Exp. Med. 180, 2287-2295, 1994
A:Title: Expression of different members of the Ly-49 gene family defines distinct natu
A:Reference number: I49114; MUID:95053763; PMID:7964501
A:Accession: I49115
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-43, 'Q', 45-267 <RES>
A:Cross-references: EMBL:U12890; NID:9602409; PIDN:AAA58705.1; PID:9602410
C:Genetics:
A:Gene: Ly49G.4
C:Superfamily: natural killer cell receptor P1; C-type lectin homology

Query Match          67.3%; Score 35; DB 2; Length 267;
Best Local Similarity 83.3%; Pred. NO. 50;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY      2 ETINCH 7
DB      80 ETLNCH 85

RESULT 9
I55686
LGL-1 - mouse
C:Species: Mus sp. (mouse)
C>Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 19-May-2000
C:Accession: I55686
R:Mason, L.H.; Ortaldo, J.R.; Young, H.A.; Kumar, V.; Bennett, M.; Anderson, S.K.
J. Exp. Med. 182, 293-303, 1995
A:Title: Cloning and functional characteristics of murine large granular lymphocyte-1: a
A:Reference number: I55686; MUID:9535829; PMID:7629495
A:Accession: I55686
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA

A:Residues: 1-267 <RES>
A:Cross-references: GB:S78689; NID:91041888; PIDN:AAC60707.1; PID:91041889
C:Genetics:
A:Gene: LGL-1
C:Superfamily: natural killer cell receptor P1; C-type lectin homology

Query Match          67.3%; Score 35; DB 2; Length 267;
Best Local Similarity 83.3%; Pred. NO. 50;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY      2 ETINCH 7
DB      80 ETLNCH 85

RESULT 10
I49052
Ly-49G.1 antigen - mouse
C:Species: Mus musculus (house mouse)
C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 19-May-2000
C:Accession: I49052
R:Smith, H.R.; Karhofer, F.M.; Yokoyama, W.M.
J. Immunol. 153, 1068-1079, 1994
A:Title: Ly-49 multigene family expressed by IL-2-activated NK cells.
A:Reference number: I49049; MUID:94300068; PMID:8027540
A:Accession: I49052
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-280 <RES>
A:Cross-references: EMBL:U10093; NID:9533493; PIDN:AAA50221.1; PID:9533494
C:Superfamily: natural killer cell receptor P1; C-type lectin homology

Query Match          67.3%; Score 35; DB 2; Length 280;
Best Local Similarity 83.3%; Pred. NO. 52;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY      2 ETINCH 7
DB      80 ETLNCH 85

RESULT 11
DB4968
hypothetical protein [imported] - Buchnera sp. (strain APS)
C:Species: Buchnera sp.
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
C:Accession: DB4968
R:Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.
Nature 407, 81-86, 2000
A:Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp
A:Reference number: AB4930; MUID:20445173; PMID:10993077
A:Accession: DB4968
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-280 <STO>
A:Cross-references: GB:AP000398; GSPDB:GN00144
A:Experimental source: strain APS
C:Genetics:
A:Gene: yjfr; BU328

Query Match          67.3%; Score 35; DB 2; Length 280;
Best Local Similarity 75.0%; Pred. NO. 52;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY      1 SETINCH 8
DB      247 SRFINCH 254

RESULT 12
F95058
hypothetical protein SP0505 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
```

C;Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001  
 C;Accession: F95058  
 R;Retellin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Held  
 on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzaple,  
 Science 293, 498-506, 2001  
 A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,  
 A;Title: Complete Sequence of a virulent isolate of Streptococcus pneumoniae.  
 A;Reference number: A95000; MUID:21357209; PMID:11463916  
 A;Accession: F95058  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1426 <KUR>  
 A;Cross-references: GB:AE005672; PIDN:AAK74663.1; PID:q14971978; GSPDB:GN00164; TIGR:SP4  
 A;Experimental source: strain TIGR4  
 C;Genetics:  
 A;Gene: SP0505

Query Match 67.3%; Score 35; DB 2; Length 426;  
 Best Local Similarity 75.0%; Pred. No. 74;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 SETINCH 8  
 DB 356 SGVINCH 363

RESULT 13  
 H97927  
 type 1 site-specific deoxyribonuclease (EC 3.1.21.3) - Streptococcus pneumoniae (strain  
 C;Species: Streptococcus pneumoniae  
 C;Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 27-Nov-2001  
 C;Accession: H97927  
 R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.; E  
 y, R.; Leblang, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M  
 J. Bacteriol. 183, 5709-5717, 2001  
 A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;  
 A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
 A;Reference number: A97872; MUID:21429245; PMID:11544234  
 A;Accession: H97927  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1522 <KUR>  
 A;Cross-references: GB:AE007317; PIDN:AAK9252.1; PID:g15458016; GSPDB:GN00174  
 C;Genetics:  
 A;Gene: hsdS  
 C;Keywords: hydrolase

Query Match 67.3%; Score 35; DB 2; Length 522;  
 Best Local Similarity 75.0%; Pred. No. 89;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 SETINCH 8  
 DB 452 SGVINCH 459

RESULT 14  
 VCLJG2  
 env polyprotein precursor - human immunodeficiency virus type 2 (isolate ROD)  
 N;Alternate names: coat polyprotein  
 C;Species: human immunodeficiency virus type 2, HIV-2  
 C;Date: 30-Sep-1987 #sequence\_revision 30-Sep-1987 #text\_change 16-Jul-1999  
 C;Accession: C26262  
 R;Guyader, M.; Emerman, M.; Sonigo, P.; Clavel, F.; Montagnier, L.; Alizon, M.  
 Nature 326, 662-669, 1987  
 A;Title: Genome organization and transactivation of the human immuno-deficiency virus ty  
 A;Reference number: A26262; MUID:87173056; PMID:3031510  
 A;Contents: proviral DNA  
 A;Accession: C26262  
 A;Molecule type: DNA  
 A;Residues: 1-858 <GUY>

A;Cross-references: GB:M15390; NID:g1332361; PIDN:AA00770.1; PID:g325749  
 C;Genetics:  
 A;Gene: env  
 C;Superfamily: type E retrovirus env polyprotein  
 C;Keywords: AIDS; capsid protein; coat protein; glycoprotein; immunodeficiency; poly  
 F;1-17/Domin: signal sequence #status predicted <SIG>  
 F;18-858/Product: env polyprotein #status predicted <MAT>  
 F;18-501/Product: exterior membrane glycoprotein #status predicted <EXT>  
 F;502-858/Product: transmembrane glycoprotein #status predicted <TM>  
 F;34,67,76,119,120,151,166,179,192,193,196,206,238,241,248,272,278,289,300,367,371,40

Query Match 67.3%; Score 35; DB 1; Length 858;  
 Best Local Similarity 85.7%; Pred. No. 14e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 ETINCH 8  
 DB 159 ETINCH 165

RESULT 15  
 T25539  
 hypothetical protein C10H11.9 - Caenorhabditis elegans  
 C;Species: Caenorhabditis elegans  
 C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 11-Jan-2000  
 C;Accession: T25539  
 R;Dante, M.; Wamsley, P.  
 submitted to the EMBL data library, February 1997  
 A;Description: The sequence of C. elegans cosmid C10H11.  
 A;Reference number: Z20047  
 A;Accession: T25539  
 A;Status: preliminary; translated from GB/EMBL/DDAB  
 A;Molecule type: DNA  
 A;Residues: 1-1173 <DAN>  
 A;Cross-references: EMBL:U08311; PIDN:AA042348.1; GSPDB:GN00019; CESP:C10H11.9  
 A;Experimental source: strain Bristol N2; clone C10H11  
 C;Genetics:  
 A;Gene: CESP:C10H11.9  
 A;Map position: 1  
 A;Introns: 23/3; 51/1; 104/3; 343/1; 478/2; 868/3; 891/3; 970/3; 1027/3; 1114/1  
 C;Superfamily: hypothetical protein C10H11.9; protein kinase homology

Query Match 67.3%; Score 35; DB 2; Length 1173;  
 Best Local Similarity 75.0%; Pred. No. 1.8e+02;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 ETINCH 9  
 DB 1091 ECKNCH 1098

Search completed: July 11, 2003, 12:00:31  
 Job time : 2.91497 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:43:41 ; Search time 0.860707 seconds

(without alignments)  
433.698 Million cell updates/sec

Title: US-10-091-442-1

Perfect score: 52

Sequence: 1 SETINCHK 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SWISSPROT\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	52	100.0	380	1 SPB7_HUMAN	O75635 homo sapien
2	37	71.2	320	1 AA3R_RAT	P28647 rattus norv
3	36	69.2	863	1 YK44_YEAST	P36023 saccharomyc
4	35	67.3	118	1 YEO6_HAEIN	P44181 haemophilus
5	35	67.3	280	1 KLR7_MOUSE	O60654 mus musculu
6	35	67.3	338	1 RSMC_BUCAL	P57413 buchnera ap
7	35	67.3	858	1 ENV_HV2RO	P04577 human immun
8	35	67.3	2095	1 RRP1_TOSV	P37800 toscana vir
9	35	67.3	2156	1 RPI_HUMAN	P56715 homo sapien
10	34	65.4	215	1 HB2L_TAMST	O06577 tamias sibi
11	34	65.4	231	1 HB2L_CHICK	P23068 gallus gall
12	34	65.4	436	1 ACHX_ONCYO	P54247 onchocerca
13	34	65.4	1078	1 CASR_HUMAN	P41180 homo sapien
14	34	65.4	1079	1 CASR_MOUSE	O94966 mus musculu
15	34	65.4	1079	1 CASR_RAT	P48442 rattus norv
16	34	65.4	1085	1 CASR_BOVIN	P35384 bos taurus
17	33	63.5	313	1 PENA_BURCE	O02940 burkholderi
18	33	63.5	319	1 AA3R_MOUSE	O64618 mus musculu
19	33	63.5	320	1 RLPA_RICPR	O94261 rickettsia
20	33	63.5	701	1 TREP_XENLA	P20033 xenopus lae
21	33	63.5	764	1 PTGR_HUMAN	P01833 homo sapien
22	32	61.5	69	1 CARP_SCHMA	P13566 schistosoma
23	32	61.5	92	1 NL7B_HUMAN	P10974 rictinus com
24	32	61.5	178	1 RL25_HELPD	O921C4 heliobacte
25	32	61.5	226	1 14E_HUMAN	O14530 homo sapien
26	32	61.5	236	1 PTL_SHEEP	P16038 ovis aries
27	32	61.5	250	1 D250_ASFB7	P32092 african swi
28	32	61.5	277	1 BHMT_PIG	O95332 sus scrofa
29	32	61.5	287	1 SCR_CHICK	O09108 gallus gall
30	32	61.5	306	1 Y361_HAEIN	P44662 haemophilus
31	32	61.5	320	1 RLPA_RICCN	O92183 rickettsia
32	32	61.5	327	1 ANX8_HUMAN	P13928 homo sapien
33	32	61.5	368	1 VGS9_HSVSA	O01052 herpesvirus

34	32	61.5	406	1 BHMT_HUMAN	O93088 homo sapien
35	32	61.5	407	1 BHMT_MOUSE	O35490 mus musculu
36	32	61.5	407	1 BHMT_RAT	O09171 rattus norv
37	32	61.5	750	1 NUSC_TRECS	O91C2 tecoma scan
38	32	61.5	763	1 PEPX_LACIC	O9CE01 lactococcus
39	32	61.5	763	1 PEPX_EEYVA	P22346 lactococcus
40	32	61.5	1254	1 POLS_EEYVA	P36331 venezuelan
41	32	61.5	1660	1 VIT6_OSCBR	O94637 oscobius br
42	32	61.5	3511	1 MY15_MOUSE	O94924 mus musculu
43	31.5	60.6	295	1 C1N5_YEAST	P40917 saccharomyc
44	31	59.6	130	1 LRG2_MOUSE	O9C9W5 mus musculu
45	31	59.6	130	1 LRG2_RAT	O92144 rattus norv

## ALIGNMENTS

RESULT 1  
ID SPB7\_HUMAN STANDARD: PRT; 380 AA.  
AC O75635;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Megsin (TP55) (Serpin B7).  
GN SERPINB7.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1].  
RX MEDLINE=97326116; PubMed=9182567;  
RA Tsurumoto M., Tsurumoto N., Ishida N., Kurihara T., Iwasa F.,  
RA Yamashiro K., Rogi T., Kodama S., Katsuragi N., Adachi M.,  
RA Katayama T., Nakao M., Yamaiichi K., Hashino J., Haruyama M., Miura K.,  
RA Nakanishi T., Nakazato H., Teramura M., Mizoguchi H., Yamaguchi N.;  
RT "Purification, cDNA cloning, and characterization of a new serpin with  
RT megakaryocyte maturation activity";  
RT J. Biol. Chem. 272:15373-15380(1997).  
RL [2].  
RN SEQUENCE FROM N.A.  
RP TISSUE=Mesangial cells;  
RC MEDLINE=98376492; PubMed=9710452;  
RA Miyata T., Nangaku M., Suzuki D., Inagi R., Uragami K., Sakai H.,  
RA Okubo K., Kurokawa K.;  
RT "A mesangium-predominant gene, megsin, is a new serpin upregulated in  
RT Iga nephropathy";  
RT J. Clin. Invest. 102:828-836(1998).  
RL  
CC -!- FUNCTION: Might function as an inhibitor of Lys-specific  
CC proteases. Might influence the maturation of megakaryocytes via  
CC its action as a serpin.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -!- TISSUE SPECIFICITY: Predominantly expressed in mesangial cells.  
CC -!- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
CC or send an email to [license@sdb.ch](mailto:license@sdb.ch)).  
CC  
CC EMBL: D88575; BAA1232.1; -  
CC EMBL: AF027866; AAC64506.1; -  
CC HSP: P05619; 1HLE.  
CC Genew: HGNC:13902; SERPINB7.  
CC MIM: 603357; -  
CC InterPro: IPR000215; Serpin.  
CC Pfam: PF00079; serpin; 1.  
CC SMART: SM00093; SERPIN; 1.  
CC PROSITE: PS00284; SERPIN; 1.  
DR

KW Serpin: Serine protease inhibitor.  
 FT ACTSITE 347 348 REACTIVE\_BOND (BY SIMILARITY).  
 SQ SEQUENCE 380 AA: 42904 MW: 9A2CDB6C63CFF605 CRC64;  
 Query Match 100.0%; Score 52; DB 1; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 0.0081;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SETINCHFRK 9  
 Db 188 SETINCHFRK 196

RESULT 2  
 ID AA3R\_RAT STANDARD: PRT: 320 AA.  
 AC P28647; Q63792;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Adenosine A3 receptor (TGPR1).  
 GN ADORA3.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxId=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RC TISSUE=Brain;  
 RX MEDLINE=92366475; PubMed=1323836;  
 RA Zhou O.Y., Li C.Y., Olan M.E., Johnson R.A., Stiles G.L., Civeilli O.;  
 RT "Molecular cloning and characterization of an adenosine receptor: the  
 RT A3 adenosine receptor.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:7432-7436(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Wistar; TISSUE=Testis;  
 RX MEDLINE=91285122; PubMed=1647979;  
 RA Meyerhof W., Mueller-Brechlin R., Richter D.;  
 RT "Molecular cloning of a novel putative G-protein coupled receptor  
 RT expressed during rat spermiogenesis.";  
 RL FEBS Lett. 284:155-160(1991).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=96196578; PubMed=8612733;  
 RA Sajjadi F.G., Boyle D.L., Domingo R.C., Firestein G.S.;  
 RT "cDNA cloning and characterization of A31, an alternatively spliced  
 RT rat A3 adenosine receptor variant.";  
 RL FEBS Lett. 382:125-129(1996).  
 CC -1- FUNCTION: RECEPTOR FOR ADENOSINE. THE ACTIVITY OF THIS RECEPTOR  
 CC IS MEDIATED BY G PROTEINS WHICH INHIBITS ADENYL CYCLASE.  
 CC POSSIBLE ROLE IN REPRODUCTION.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS: TWO ISOFORMS ARE PRODUCED BY ALTERNATIVE  
 CC SPLICING.  
 CC -1- TISSUE SPECIFICITY: TESTIS, PARTICULARLY IN SPERMATOCYTES AND  
 CC SPERMATIDS BUT NOT IN SPERMATOGENIA. LOW LEVELS IN THE BRAIN.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING SPERMIOGENESIS.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: X59249; CAA41937.1; -  
 CC DR EMBL: M94152; AAA40680.1; -  
 CC DR EMBL: X93219; CAA63702.1; -  
 CC PIR: S17177; S17177.

DR PIR: A46152; A46152.  
 DR HSSP: P29274; 1MMH.  
 DR InterPro: IPR000276; GPCR\_Rhodpsn.  
 DR Pfam: PF00001; 7tm.1; 1.  
 DR PRINTS: PR00237; GPCRHHODPSN.  
 DR PROSITE: PS00237; G-PROTEIN\_RECPT\_FL1; 1.  
 DR PROSITE: PS50262; G-PROTEIN\_RECPT\_FL2; 1.  
 KW G-protein coupled receptor; Transmembrane; Glycoprotein;  
 KW Lipoprotein; Palmitate; Alternative splicing.  
 FT DOMAIN 1 16  
 FT TRANSMEM 17 39  
 FT DOMAIN 40 50  
 FT TRANSMEM 51 74  
 FT DOMAIN 75 86  
 FT TRANSMEM 87 108  
 FT DOMAIN 109 128  
 FT TRANSMEM 129 150  
 FT DOMAIN 151 179  
 FT TRANSMEM 180 200  
 FT DOMAIN 201 233  
 FT TRANSMEM 234 257  
 FT DOMAIN 258 263  
 FT TRANSMEM 264 286  
 FT DOMAIN 287 320  
 FT CARBOHYD 4 4  
 FT CARBOHYD 5 5  
 FT DISULFID 85 168  
 FT LIPID 305 305  
 FT VARSPPLIC 119 119  
 FT CONFLICT 18 18  
 FT CONFLICT 68 68  
 FT CONFLICT 74 83  
 FT CONFLICT 132 132  
 FT CONFLICT 172 172  
 FT CONFLICT 288 288  
 FT CONFLICT 291 295  
 SQ SEQUENCE 320 AA: 36629 MW: 59DBA40C594ACDAB CRC64;  
 Query Match 71.2%; Score 37; DB 1; Length 320;  
 Best Local Similarity 55.6%; Pred. No. 5.3;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 SETINCHFRK 9  
 Db 163 SSTLSCHFR 171

RESULT 3  
 ID YK44\_YEAST STANDARD: PRT: 863 AA.  
 AC P36023;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Putative 101.8 kDa transcriptional regulatory protein in *IAS1-CCPI*  
 DE intergenic region.  
 GN YKR064W.  
 OS *Saccharomyces cerevisiae* (Baker's yeast).  
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxId=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288c;  
 RA van Vliet-Reedijk J.C., Planta R.J.;  
 RU Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -1- SIMILARITY: CONTAINS 1 ZN(2)-CYS(6), FUNGAL-TYPE BINUCLEAR  
 CC CLUSTER DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its



CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; 228289; CAAB2143.1; -  
DR PIR; S38140; S38140.  
DR HSSP; P12351; 1HWT.  
DR SGD; S0001772; YKR064W.  
DR InterPro; IPR001138; Fung1\_Trn.  
DR Pfam; PF00172; Zn\_C1us; 1.  
DR SMART; SM00066; GAL4; 1.  
DR PROSITE; PS00463; ZN2\_Cy6\_FUNGAL\_1; 1.  
DR PROSITE; PS00448; ZN2\_Cy6\_FUNGAL\_2; 1.  
KW Hypothetical protein: Transcription regulation; DNA-binding;  
KW Nuclear protein; zinc; metal-binding.  
FT DNABIND 18 47  
SO SEQUENCE 863 AA; 101824 MW; B140465B405FAF84 CRC64;

Query Match 69.2%; Score 36; DB 1; Length 863;  
Best Local Similarity 62.5%; Pred. No. 21;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 ETINCHER 9  
DB 470 KTLCHFK 477

RESULT 4  
YE06\_HAEIN STANDARD: PRT; 118 AA.  
ID YE06\_HAEIN

AC P44181;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein H11406.  
GN H11406.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_Taxid:727;

RP SEQUENCE FROM N.A.  
RC STRAIN-Rd / KM20 / ATCC 51907;  
RX MEDLINE-95350630; PubMed-7542800;  
RA Fletschmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Uutterback T.R., Hanna M.C., Nguyen D.T., Sauder D.W., Brandon R.C.,  
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Georgiagen N.S.M.,  
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
Rd.";  
RL Science 269:496-512(1995).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; U32820; AAC23056.1; -  
DR TIGR; H11406; -  
KW Hypothetical protein: Complete proteome.  
SO SEQUENCE 118 AA; 13325 MW; C1CAF33E6CB83452 CRC64;

Query Match 67.3%; Score 35; DB 1; Length 118;  
Best Local Similarity 66.7%; Pred. No. 5;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
OY 1 SETINCHER 9  
DB 35 SETSYHFK 43

RESULT 5  
KRL7\_MOUSE  
ID KRL7\_MOUSE STANDARD: PRT; 280 AA.  
AC Q60654; Q60655; Q60683;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Killer cell lectin-like receptor 7 (T-cell surface glycoprotein  
DE LY-49G) (LY49-G antigen).  
GN KLR7 OR LY49G OR LY-49G OR LY49-G OR LY49G4.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid:10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6; Tissue-Spleen;  
RX MEDLINE-94300068; PubMed-8027540;  
RA Smith H.R.C., Karlhofer F.M., Yokoyama W.M.;  
RT "Ly-49 multigene family expressed by IL-2-activated NK cells.";  
RL J. Immunol. 153:1068-1079(1994).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM LY-49G.2).  
RC STRAIN-B10.A;  
RX MEDLINE-95053763; PubMed-7964501;  
RA Brennan J., Mager D., Jefferies W., Takel F.;  
RT "Expression of different members of the Ly-49 gene family defines  
RT distinct natural killer cell subsets and cell adhesion properties.";  
RL J. Exp. Med. 180:2287-2295(1994).  
CC -1- FUNCTION: RECEPTOR ON NATURAL KILLER (NK) CELLS FOR CLASS I MHC.  
CC -1- SUBUNIT: HOMODIMER. DISULFIDE-LINKED.  
CC -1- SUBCELLULAR LOCATION: Type II membrane protein.  
CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS: LY-49G.1 (SHOWN HERE), LY-  
CC 49G.2 AND LY-49G.3; ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; U10093; AAA50221.1; -  
DR EMBL; U10094; AAA50222.1; -  
DR EMBL; U10095; AAA50223.1; -  
DR EMBL; U12890; AAA58705.1; -  
DR MGD; MGI:101901; Klr7.  
DR InterPro; IPR001304; Lectin\_C.  
DR Pfam; PF00059; Lectin\_C; 1.  
DR SMART; SM00034; CLECT; 1.  
DR PROSITE; PS00615; C-TYPE\_LECTIN\_1; FALSE\_NEG.  
DR PROSITE; PS00441; C-TYPE\_LECTIN\_2; 1.  
KW T-cell; Glycoprotein; Antigen; Transmembrane; Cell adhesion;  
KW Signal-anchor; Lectin; Receptor; Multigene family;  
KW Alternative splicing.  
FT DOMAIN 1 44  
FT TRANSMEM 45 66  
FT DOMAIN 67 280  
FT DOMAIN 156 275  
FT DISULFID 185 271  
FT DISULFID 250 263  
FT CARBOHYD 87 87  
FT CARBOHYD 104 104

CYTOPLASMIC (POTENTIAL).  
SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
(POTENTIAL).  
EXTRACELLULAR (POTENTIAL).  
C-TYPE LECTIN (LONG FORM).  
BY SIMILARITY.  
BY SIMILARITY.  
N-LINKED (GLCNAC. . .) (POTENTIAL).  
N-LINKED (GLCNAC. . .) (POTENTIAL).

```

FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VASAPLIC 117 193 MISSING (IN ISOFORM LY-49G3).
FT VASAPLIC 142 154 MISSING (IN ISOFORM LY-49G2).
FT CONFLICT 44 44 K -> O (IN REF. 2).
SQ SEQUENCE 280 AA; 32522 MW; 4704A2D87D5E83FB CRC64;

Query Match 67.3%; Score 35; DB 1; Length 280;
Best Local Similarity 83.3%; Pred. No. 11;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 ETINCH 7
Db 80 ETINCH 85

RESULT 6
RSMC_BUCAL STANDARD; PRT: 338 AA.
ID p57413;
AC p57413; 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ribosomal RNA small subunit methyltransferase C (EC 2.1.1.52) (rRNA
DE (guanine-N(2)-methyltransferase) (16S rRNA m2G1207
DE methyltransferase).
GN RSMC OR B0328.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Tokyo 1998; PubMed=10993077;
RA MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Matanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. Aph. ";
RL Nature 407:81-86(2000).
CC -1- FUNCTION: SPECIFICALLY METHYLATEDS THE GUANOSINE IN POSITION 1207
CC OF 16S RNA IN THE 30S PARTICLE (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + RNA = S-adenosyl-L-
CC homocysteine + RNA containing N(2)-methylguanine.
CC -1- SIMILARITY: BELONGS TO THE METHYLTRANSFERASE SUPERFAMILY. RSMC
CC SUPERFAMILY.
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A POTENTIAL FRAMESHIFT
CC WAS CORRECTED IN POSITION 57.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, AP001119; BAB13036.1; ALT_FRAME.
DR InterPro: IPR002052; N6_Mtase.
DR InterPro: IPR000051; SAM_Bind.
DR PROSITE: PS00092; N6_MTASE; UNKNOWN.1.
KW rRNA processing; Transferase; Methyltransferase; Complete proteome.
SQ SEQUENCE 338 AA; 39105 MW; E43516DD22FA014 CRC64;

Query Match 67.3%; Score 35; DB 1; Length 338;
Best Local Similarity 75.0%; Pred. No. 13;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 SETINCH 8
Db 305 SREFINCH 312

RESULT 7
ENV_HVZRO

```

```

ID ENV_HVZRO STANDARD; PRT: 858 AA.
AC P04577;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Envelope glycoprotein gp160 precursor [contains: Exterior membrane
DE glycoprotein (gp120); Transmembrane glycoprotein (gp41)].
GN ENV.
OS Human immunodeficiency virus type 2 (isolate ROD) (HIV-2).
OC Viruses; Retroviruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11720;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87173056; PubMed=3031510;
RA Guyader M., Emerman M., Sonigo P., Clavel F., Montagnier L.,
RA Alizon M.;
RT "Genome organization and transactivation of the human
RT immunodeficiency virus type 2.";
RL Nature 326:662-669(1987).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, M15390; AAB00770.1; -
DR EMBL, X05291; CAA28914.1; -
DR PIR, C26282; VCLG2.
DR HIV, M15390; ENV52ROD.
DR InterPro: IPR000328; Env_GP41.
DR InterPro: IPR000777; GP120.
DR Pfam, PF00516; GP120; 1.
DR Pfam, PF00517; GP41; 1.
KW AIDS; Coat protein; Glycoprotein; Transmembrane;
KW Signal.
KW CHAIN
FT 1 17
FT 18 501
FT 502 858
FT CARBOHYD 34 34 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 67 67 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 76 76 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 119 119 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 120 120 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 151 151 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 166 166 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 179 179 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 192 192 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 193 193 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 206 206 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 238 238 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 241 241 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 248 248 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 272 272 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 278 278 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 289 289 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 310 310 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 371 371 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 371 371 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 400 400 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 447 447 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 463 463 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 466 466 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 611 611 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 636 636 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 636 636 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 312 312 T -> I (IN REF. 1; AAB00770).
SQ SEQUENCE 858 AA; 98824 MW; C7266AFLF5C5B9A7 CRC64;

```

Query Match 67.3%; Score 35; DB 1; Length 858;  
 Best Local Similarity 85.7%; Pred No. 32;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 ETINCHP 8  
 DB 159 ETINCOF 165

RESULT 9  
 RPL\_TOSV STANDARD; PRT: 2095 AA.  
 AC P37800;  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE RNA-directed RNA polymerase (EC 2.7.7.48) (L protein).  
 GN L.  
 OS Toscana virus (Tos).  
 OC Viruses; ssRNA negative-strand viruses; Bunyaviridae; Phlebovirus.  
 OX NCBI\_TaxID=11590;  
 RN 11;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ISS.PHL.3;  
 RX MEDLINE=93312587; PubMed=8460526;  
 RA Accardi L., Gro M.C., di Bonito P., Giorgi C.;  
 RT "Toscana virus genomic L segment: molecular cloning, coding strategy  
 RT and amino acid sequence in comparison with other negative strand RNA  
 RT viruses.";  
 RL Virus Res. 27:119-131(1993).  
 CC -1 CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +  
 CC (RNA)(N).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL: X68414; CAA48478.1; -  
 CC DR PIR: S29529; S29529.  
 CC TRANSFERRASE: Nucleotidyltransferase; RNA-directed RNA polymerase.  
 CC KM SEQUENCE 2095 AA; 238885 MW; 22FF4DAD745583B7 CRC64;  
 CC SO

Query Match 67.3%; Score 35; DB 1; Length 2095;  
 Best Local Similarity 62.5%; Pred. No. 74;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 ETINCHP 9  
 DB 501 ETLDCHLK 508

RESULT 9  
 RPL\_HUMAN STANDARD; PRT: 2156 AA.  
 AC P56715;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Oxygen-regulated protein 1 (Retinitis pigmentosa Rpl protein)  
 DE (Retinitis pigmentosa 1 protein).  
 GN RPL1 OR ORP1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN 11;  
 RP SEQUENCE FROM N.A., AND VARIANTS HIS-872; TYR-985; THR-1670; PRO-1691  
 RP AND TYR-2033.

RC TISSUE=Retina;  
 RX MEDLINE=93318096; PubMed=10391212;  
 RA Sullivan L.S., Heckenlively J.R., Byrne S.J., Zuo J., Hyde M.A.,  
 RA Gal A., Denton M., Inglehearn C.F., Blanton S.H., Daiger S.P.;  
 RT "Mutations in a novel retina-specific gene cause autosomal dominant  
 RT retinitis pigmentosa.";  
 RL Nat. Genet. 22:255-259(1999).  
 RN 12;  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Retina;  
 RX MEDLINE=93318095; PubMed=10391211;  
 RA Pierce E.A., Quinn T., Mehan T., McGee T.L., Berson E.L., Dryja T.P.;  
 RT "Mutations in a gene encoding a new oxygen-regulated photoreceptor  
 RT protein cause dominant retinitis pigmentosa.";  
 RL Nat. Genet. 22:248-254(1999).  
 RN 13;  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99330563; PubMed=10401003;  
 RA Guillonnet X., Pirliev N.I., Danciger M., Kozak C.A., Clidecyan A.V.,  
 RA Jacobson S.G., Farber D.B.;  
 RT "A nonsense mutation in a novel gene is associated with retinitis  
 RT pigmentosa in a family linked to the RPL locus.";  
 RL Hum. Mol. Genet. 8:1541-1546(1999).  
 CC -1 FUNCTION: COULD HAVE A ROLE IN THE DIFFERENTIATION OF  
 CC PHOTORECEPTOR CELLS.  
 CC -1 TISSUE SPECIFICITY: EXPRESSED IN RETINA, NOT EXPRESSED IN HEART,  
 CC BRAIN, PLACENTA, LUNG, LIVER, SKELETAL MUSCLE, KIDNEY, SPLEEN AND  
 CC PANCREAS.  
 CC -1 DISEASE: DEFECTS IN RPL CAUSE RETINITIS PIGMENTOSA FORM 1 (RPL1); A  
 CC DISEASE CHARACTERIZED BY CONSTRUCTION OF THE VISUAL FIELDS, NIGHT  
 CC BLINDNESS, AND FUNDUS CHANGES. THE DISEASE SEEMS TO BE ASSOCIATED  
 CC WITH TRUNCATED (STOP OR FRAMESHIFT MUTATIONS) FORMS OF THE  
 CC PROTEIN.  
 CC -1 SIMILARITY: CONTAINS 2 DOUBLECORTIN DOMAINS.  
 CC -1 DATABASE: NAME=RetNet;  
 CC NOTE=Retinal information network;  
 CC WWW="http://www.sph.utmc.edu/retnet/".  
 CC -1 DATABASE: NAME=Mutations of the RPL gene;  
 CC NOTE=Retina International's Scientific Newsletter;  
 CC WWW="http://www.retina-international.com/sci-news/rplmut.htm".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL: AF143226; AAD44197.1; -  
 CC DR EMBL: AF143224; AAD44197.1; JOINED.  
 CC DR EMBL: AF143225; AAD44197.1; JOINED.  
 CC DR EMBL: AF143222; AAD44198.1; -  
 CC DR EMBL: AF141021; AAD42072.1; -  
 CC DR EMBL: AF152242; AAD46774.1; -  
 CC DR EMBL: AF152240; AAD46774.1; JOINED.  
 CC DR EMBL: AF152241; AAD46774.1; JOINED.  
 CC DR EMBL: AF145592; AAD46769.1; -  
 CC DR GeneW: HGNC:10263; RPL1.  
 CC DR MIM: 603937; -  
 CC DR MIM: 180100; -  
 CC DR InterPro: IPR003533; DCX.  
 CC DR Pfam: PF03607; DCX; 2.  
 CC DR SMART: SM00537; DCX; 2.  
 CC DR PROSITE: PS50309; DCX; 2.  
 CC KW Vision; Retinitis pigmentosa; Polymorphism; Repeat.  
 CC FT DOMAIN 36 118 DOUBLECORTIN 1.  
 CC FT FT 154 233 DOUBLECORTIN 2.  
 CC FT DOMAIN 268 273 POLY-SER.  
 CC FT DOMAIN 671 675 POLY-LYS.  
 CC FT DOMAIN 1687 1691 POLY-SER.  
 CC FT VARIANT 872 872 R -> H.  
 CC /FTID=VAR\_007810.

```

FT VARIANT 985 985 N->Y. /FTID=VAR_007811.
FT VARIANT 1670 1670 A->T. /FTID=VAR_007812.
FT VARIANT 1691 1691 S->P. /FTID=VAR_007813.
FT VARIANT 2033 2033 C->Y. /FTID=VAR_007814.
SQ SEQUENCE 2156 AA; 240659 MW; 55ADEBC43D6A507 CRC64;

Query Match
Best Local Similarity 67.3%; Score 35; DB 1; Length 2156;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 ETINCHP 8
Db 2089 ENINCHY 2095

RESULT 10
HP27_TAMSI STANDARD; PRT; 215 AA.
AC 006577;
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hibernation-associated plasma protein HP-27 precursor (hibernator-
specific blood complex, 27 kDa subunit).
OS Tamias sibiricus (Siberian chipmunk) (Asian chipmunk).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Sciuridae; Sciurinae;
OC Tamias.
OX NCBI_TaxID=64680;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=93180798; PubMed=8441393;
RA Takamatsu N., Oba K., Kondo J., Kondo N., Shiba T.;
RT "Hibernation-associated gene regulation of plasma proteins with a
collagen-like domain in mammalian hibernators.";
RL Mol. Cell. Biol. 13:1516-1521(1993).
RN [2]
RP SEQUENCE OF 31-215.
RC TISSUE=Plasma;
RX MEDLINE=92112696; PubMed=1730610;
RA Kondo N., Kondo J.;
RT "Identification of novel blood proteins specific for mammalian
hibernation.";
RL J. Biol. Chem. 267:473-478(1992).
CC -1- FUNCTION: PLASMA PROTEINS HP-20, HP-25, HP-27 AND HP-55 FORM A
140 kDa COMPLEX VIA DISULFIDE BONDS IN THE PLASMA AND ARE
HIBERNATION SPECIFIC.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: PLASMA; SYNTHESIZED IN THE LIVER.
CC -1- DEVELOPMENTAL STAGE: THE PROTEIN COMPLEX DISAPPEARS FROM THE
PLASMA AT ONSET OF HIBERNATION AND REAPPEARS AS HIBERNATION
CEASES.
CC -1- SIMILARITY: CONTAINS 1 COLLAGENOUS DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 C10 DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D12976; BAA02353.1; -
CC InterPro; IPR001073; C1Q.
CC InterPro; IPR000087; Collagen.
CC Pfam; PF00386; C1q; 1.
CC Pfam; PF01391; Collagen; 1.

```

```

DR PRINTS; PR00007; COMPLEMENTC1Q.
DR SMART; SM00110; C1Q; 1.
DR PROSITE; PS01113; C1Q; 1.
KW Signal; collagen; glycoprotein; plasma; Multigene family.
FT SIGNAL 1 30
FT CHAIN 31 215
FT DOMAIN 43 81
FT DOMAIN 83 215
FT CARBOHYD 155 155
FT CONFLICT 41 41
FT CONFLICT 51 51
SQ SEQUENCE 215 AA; 22797 MW; D2357086A088FA9E CRC64;

Query Match
Best Local Similarity 71.4%; Score 34; DB 1; Length 215;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 3 TINCHP 9
Db 81 TVNCHSK 87

RESULT 11
HB2L_CHICK STANDARD; PRT; 231 AA.
ID HB2L_CHICK
AC P23068;
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-NOV-1991 (Rel. 20, Last annotation update)
DE Class II histocompatibility antigen, B-L beta chain (Fragment).
DE Gallus gallus (Chicken).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A. (CLONE p14).
RX MEDLINE=88296406; PubMed=2841107;
RA Bourlet Y., Behar G., Guillemot F., Frechin N., Billault A.,
Chausse A.M., Zoorob R., Auffray C.;
RT "Isolation of chicken major histocompatibility complex class II (B-L)
beta chain sequences: comparison with mammalian beta chains and
expression in lymphoid organs.";
RL EMBO J. 7:1031-1039(1988).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X07447; -; NOT_ANNOTATED_CDS.
DR PIR; S00475; H1CHBL.
DR HSSP; P01888; IBMG.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR000353; MHC_II_beta.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00969; MHC_II_beta; 1.
DR ProDom; PD000328; MHC_II_beta; 1.
DR SMART; SM00407; IGc1; 1.
DR PROSITE; PS00290; IG_MHC; 1.
KW MHC II; Transmembrane; Glycoprotein.
FT NON_TER 1
FT DOMAIN 89 1
FT DOMAIN 90 182
FT DOMAIN 183 194
FT TRANSMEM 195 219
FT DOMAIN 220 231
FT DISULFID 10 74
CYTOPLASMIC TAIL.
BY SIMILARITY.

```

FT DISULFID 111 167 BY SIMILARITY.  
 FT CARBOHYD 14 14 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SO SEQUENCE 231 AA; 26278 MW; 06ADE955FAFB22A CRC64;  
 Query Match 65.4%; Score 34; DB 1; Length 231;  
 Best Local Similarity 62.5%; Pred. No. 15;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 OY 1 SETINCHP 8  
 DB 5 SATVECHF 12  
 RESULT 12  
 ACHX\_ONCVO .STANDARD; PRT; 436 AA.  
 AC P54247.  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Acetylcholine receptor protein, non-alpha chain (Fragment).  
 OS Onchocerca volvulus.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Splirurida; Filarioidae;  
 OC Onchocercidae; Onchocerca.  
 OX NCBI\_TaxID=6282;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94299155; PubMed=8026747;  
 RA Ajun P.M., Eswang T.G.;  
 RT "Cloning of a cDNA encoding a putative nicotinic acetylcholine  
 RT receptor subunit of the human filarial parasite Onchocerca  
 RT volvulus.";  
 RL Gene 144:127-129(1994).  
 CC -1- FUNCTION: AFTER BINDING ACETYLCHOLINE, THE ACHR RESPONDS BY AN  
 CC EXTENSIVE CHANGE IN CONFORMATION THAT AFFECTS ALL SUBUNITS AND  
 CC LEADS TO OPENING OF AN ION-CONDUCTING CHANNEL ACROSS THE PLASMA  
 CC MEMBRANE.  
 CC -1- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNEL FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC -----  
 CC DR EMBL; L20465; AAA21823.1; .  
 CC DR EMBL; L12543; AAA29415.1; .  
 CC DR InterPro: IPR000188; GABA\_A\_receptor.  
 CC DR InterPro: IPR001175; Neur\_Channel.  
 CC Pfam: PF02931; Neur\_chan\_LBD; 1.  
 CC Pfam: PF02932; Neur\_chan\_memb; 1.  
 CC DR PROSITE: PS00236; NEUROTRION\_CHANNEL; 1.  
 CC KM Postsynaptic membrane; Ionic channel; Glycoprotein; Transmembrane;  
 CC Multigene family.  
 FT NON\_TER 1  
 FT DOMAIN 1  
 FT TRANSMEM 195  
 FT TRANSMEM 196  
 FT TRANSMEM 219  
 FT TRANSMEM 227  
 FT TRANSMEM 245  
 FT TRANSMEM 261  
 FT TRANSMEM 280  
 FT DOMAIN 281  
 FT TRANSMEM 404  
 FT TRANSMEM 405  
 FT TRANSMEM 423  
 FT DISULFID 89  
 FT CARBOHYD 62  
 FT CARBOHYD 62  
 FT CARBOHYD 140  
 FT CARBOHYD 140  
 SO SEQUENCE 436 AA; 51340 MW; 58051BFES9A4555F CRC64;  
 Query Match 65.4%; Score 34; DB 1; Length 436;  
 Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 NCHFR 9  
 DB 102 NCHFR 106  
 RESULT 13  
 CASR\_HUMAN .STANDARD; PRT; 1078 AA.  
 ID CASR\_HUMAN  
 AC P41180; Q13912; Q16379; Q16108; Q16109; Q16110;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Extracellular calcium-sensing receptor precursor (Casr) (Parathyroid  
 DE cell calcium-sensing receptor).  
 GN CASR OR GPRC2A OR PCARL.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX Pearce S.H.S., Thakker R.V.;  
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC Tissue-Parathyroid;  
 RX MEDLINE=95279439; PubMed=7759551;  
 RA Garrett J.E., Capuano I.V., Hammerland L.G., Hung B.C., Brown E.M.,  
 RA Hebert S.C., Nemeth E.F., Fuller F.;  
 RT "Molecular cloning and functional expression of human parathyroid  
 RT calcium receptor cDNAs.";  
 RL J. Biol. Chem. 270:12919-12925(1995).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC Tissue-Kidney;  
 RX MEDLINE=95408281; PubMed=7677761;  
 RA Aida K., Koishi S., Tawata M., Onaya T.;  
 RT "Molecular cloning of a putative Ca(2+)-sensing receptor cDNA from  
 RT human kidney.";  
 RL Biochem. Biophys. Res. Commun. 214:524-529(1995).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96343808; PubMed=8756555;  
 RA Fretschel M., Zink-Lorenz A., Hollloschl A., Hafner M., Flockerzi V.,  
 RA Raue F.;  
 RT "Expression of a calcium-sensing receptor in a human medullary  
 RT thyroid carcinoma cell line and its contribution to calcitonin  
 RT secretion.";  
 RL Endocrinology 137:3842-3848(1996).  
 RN [5]  
 RP SEQUENCE OF 643-908 FROM N.A.  
 RX MEDLINE=96193893; PubMed=8613532;  
 RA Bikle D.D., Ratnam A., Mauro T., Harris J., Pillai S.;  
 RT "Changes in calcium responsiveness and handling during keratinocyte  
 RT differentiation. Potential role of the calcium receptor.";  
 RL J. Clin. Invest. 97:1085-1093(1996).  
 RN [6]  
 RP VARIANTS FHH GLN-185; LYS-297 AND TRP-795.  
 RX MEDLINE=94094324; PubMed=7916660;  
 RA Pollak M.R., Brown E.M., Chou Y.-H.W., Hebert S.C., Marx S.J.,  
 RA Steinmann B., Levy T., Seidman C.E., Seidman J.G.;  
 RT "Mutations in the human Ca(2+)-sensing receptor gene cause familial  
 RT hypocalcaemic hypercalcaemia and neonatal severe  
 RT hyperparathyroidism.";  
 RL Cell 75:1297-1303(1993).  
 RN [7]  
 RP VARIANT ADH ALA-127.  
 RX MEDLINE=95179179; PubMed=7874174;  
 RA Pollak M.R., Brown E.M., Estep H.L., McLaine P.N., Kifor O., Park J.,  
 RA Hebert S.C., Seidman C.E., Seidman J.G.;  
 RT "Autosomal dominant hypocalcaemia caused by a Ca(2+)-sensing receptor  
 RT gene mutation.";  
 RL Nat. Genet. 8:303-307(1994).

[8]  
 RP VARIANTS FHH MET-62; CYS-66; MET-138; GLU-143 AND GLN-227.  
 RX MEDLINE-95243222; PubMed-7726161;  
 RA Chou Y.-H.M., Pollak M.R., Brandi M.L., Toss G., Arnyvist H.,  
 RA Atkinson A.B., Papapoulos S.E., Marx S., Brown E.M., Seldman J.G.,  
 RA Seldman C.E.;  
 RA "Mutations in the human Ca(2+)-sensing-receptor gene that cause  
 RT familial hypocalcemic hypercalcaemia";  
 RL Am. J. Hum. Genet. 56:1075-1079(1995).  
 RN [9]  
 RP SEQUENCE OF 1-61 FROM N.A., AND VARIANT FHH ALA-39.  
 RX MEDLINE-95403641; PubMed-7673400.  
 RA Aida K., Koishi S., Inoue M., Nakazato M., Tawata M., Onaya T.;  
 RT "Familial hypocalcemic hypercalcaemia associated with mutation in the  
 RL human Ca(2+)-sensing receptor gene";  
 RN J. Clin. Endocrinol. Metab. 80:2594-2598(1995).  
 RP VARIANTS NSHPT LEU-227 AND TYR-598.  
 RX MEDLINE-96292293; PubMed-8675635;  
 RA Pearce S.H.S., Trump D., Wooding C., Besser G.M., Chew S.L.,  
 RA Grant D.B., Heath D.A., Hughes I.A., Paterson C.R., Whyte M.P.,  
 RA Thakker R.V.;  
 RT "Calcium-sensing receptor mutations in familial benign hypercalcaemia  
 RT and neonatal hyperparathyroidism";  
 RL J. Clin. Invest. 96:2683-2692(1995).  
 RN [11]  
 RP VARIANTS ADHP THR-116; HIS-681 AND SER-806, AND VARIANT SER-851.  
 RX MEDLINE-96311554; PubMed-8733126;  
 RA Baron J., Miner K.K., Yanovski J.A., Cunningham A.W., Iane L.,  
 RA Zimmerman D., Cutler G.B. Jr.;  
 RT "Mutations in the Ca(2+)-sensing receptor gene cause autosomal  
 RT dominant and sporadic hyperparathyroidism";  
 RL Hum. Mol. Genet. 5:601-606(1996).  
 RN [12]  
 RP VARIANT FHH ARG-174.  
 RX MEDLINE-97442275; PubMed-9298824;  
 RA Ward B.K., Stuckey B.G.A., Gutteridge D.H., Laling N.G., Pullan P.T.,  
 RA Ratajczak T.;  
 RT "A novel mutation (L174R) in the Ca2+-sensing receptor gene  
 RT associated with familial hypocalcemic hypercalcaemia";  
 RL Hum. Mutat. 10:233-235(1997).  
 RN [13]  
 RP VARIANT FHH GLU-557.  
 RX MEDLINE-21603857; PubMed-11762699;  
 RA Nakayama T., Minato M., Nakagawa M., Soma M., Tobe H., Aoi N.,  
 RA Kosuge K., Sato M., Ozawa Y., Kamatsuse K., Kokubun S.;  
 RT "A novel mutation in Ca2+-sensing receptor gene in familial  
 RT hypocalcemic hypercalcaemia";  
 RL Endocrine 15:277-282(2001).  
 CC -1- FUNCTION: SENSE CHANGES IN THE EXTRACELLULAR CONCENTRATION OF  
 CC CALCIUM IONS. THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY A  
 CC G-PROTEIN THAT ACTIVATES A PHOSPHATIDYLINOSITOL-CALCIUM SECOND  
 CC MESSENGER SYSTEM.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; are  
 CC produced by alternative splicing.  
 CC -1- TISSUE SPECIFICITY: FOUND IN KIDNEY, BUT NOT IN BRAIN, LUNG,  
 CC LIVER, HEART, SKELETAL MUSCLE, OR PLACENTA.  
 CC -1- DISEASE: DEFECTS IN CASR ARE A CAUSE OF FAMILIAL HYPOCALCAEMIC  
 CC HYPERCALCAEMIA (FHH) AND NEONATAL SEVERE HYPERPARATHYROIDISM  
 CC (NSHPT). TWO INHERITED CONDITIONS CHARACTERIZED BY ALTERED CALCIUM  
 CC HOMEOSTASIS. THE MUTATIONS REDUCE THE ACTIVITY OF THE RECEPTOR.  
 CC FHH AFFECTED INDIVIDUALS EXHIBIT MILD OR MODEST HYPERCALCAEMIA,  
 CC RELATIVE HYPOCALCAEMIA, AND INAPPROPRIATELY NORMAL PTH LEVELS. IN  
 CC CONTRAST, NSHPT IS A RARE AUTOSOMAL RECESSIVE LIFE-THREATENING  
 CC DISORDER CHARACTERIZED BY VERY HIGH SERUM CALCIUM CONCENTRATIONS,  
 CC SKELETAL DEMINERALIZATION, AND PARATHYROID HYPERPLASIA. IN SOME  
 CC INSTANCES NSHPT HAS BEEN DEMONSTRATED TO BE THE HOMOZYGOUS FORM OF  
 CC FHH.  
 CC -1- DISEASE: DEFECTS IN CASR ARE ALSO THE CAUSE OF AUTOSOMAL DOMINANT  
 CC HYPOCALCAEMIA (ADH) IN WHICH THE RECEPTOR IS ACTIVATED AT SUBNORMAL  
 CC CA(2+) LEVELS.  
 CC -1- DISEASE: DEFECTS IN CASR ARE ALSO THE CAUSE OF AUTOSOMAL DOMINANT

CC HYPOPARATHYROIDISM (ADHP). ADHP IS CHARACTERIZED BY HYPOCALCAEMIA  
 CC AND HYPERPHOSPHATEMIA DUE TO INADEQUATE SECRETION OF PARATHYROID  
 CC HORMONE. SYMPTOMS ARE SEIZURES, TETANY AND CRAMPS.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X81086; CAAS6990.1; -  
 CC DR EMBL: U20759; AAA86503.1; -  
 CC DR EMBL: U20760; AAA86504.1; -  
 CC DR EMBL: D50855; BAA09453.1; -  
 CC DR EMBL: S83176; AAB46873.1; -  
 CC DR EMBL: S79217; AAB35262.2; -  
 CC DR EMBL: S68032; AAB29413.2; ALT\_SEQ.  
 CC DR EMBL: S68033; AAB29414.1; -  
 CC DR EMBL: S68036; AAB29415.1; -  
 CC DR EMBL: S81735; AAD14370.1; -  
 CC DR Genew: HGNC:1514; CASR.  
 CC DR MIM: 601199; -  
 CC DR MIM: 145980; -  
 CC DR MIM: 239200; -  
 CC DR MIM: 601198; -  
 CC DR InterPro: IPR001828; ANF\_receptor.  
 CC DR InterPro: IPR000337; GPCR\_Mgr.  
 CC DR Pfam: PF00003; 7tm\_3; 1.  
 CC DR Pfam: PF01094; ANF\_receptor; 1.  
 CC DR PRINTS: PR00248; GPCRMR.  
 CC DR PROSITE: PS00979; G-PROTEIN\_RECEP\_F3\_1; 1.  
 CC DR PROSITE: PS00980; G-PROTEIN\_RECEP\_F3\_2; 1.  
 CC DR PROSITE: PS00981; G-PROTEIN\_RECEP\_F3\_3; 1.  
 CC DR PROSITE: PS50259; G-PROTEIN\_RECEP\_F3\_4; 1.  
 CC DR KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;  
 CC Disease mutation; Alternative splicing; Polymorphism.  
 CC FT SIGNAL 1 19  
 CC FT CHAIN 20 1078  
 CC FT DOMAIN 20 612  
 CC FT TRANSMEM 613 635  
 CC FT DOMAIN 636 649  
 CC FT TRANSMEM 650 670  
 CC FT DOMAIN 671 681  
 CC FT TRANSMEM 682 700  
 CC FT DOMAIN 701 724  
 CC FT TRANSMEM 725 745  
 CC FT DOMAIN 746 769  
 CC FT TRANSMEM 770 792  
 CC FT DOMAIN 793 805  
 CC FT TRANSMEM 806 828  
 CC FT DOMAIN 829 836  
 CC FT TRANSMEM 837 862  
 CC FT DOMAIN 863 1078  
 CC FT CARBOHYD 90 90  
 CC FT CARBOHYD 130 130  
 CC FT CARBOHYD 261 261  
 CC FT CARBOHYD 287 287  
 CC FT CARBOHYD 386 386  
 CC Query Match. 65.4%; Score 34; DB 1; Length 1078;  
 CC Best local Similarity 62.5%; Pred. No. 62;  
 CC Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 CC QY 2 ETINCHK 9  
 CC II III;  
 CC Db 354 ETFNCHLQ 361  
 CC RESULT 14  
 CC CASR\_MOUSE

ID CASR\_MOUSE STANDARD: PRT; 1079 AA.  
 AC 090966; 008968; 088519; 090955; 090958; 091912;  
 DT 16-OCT-2001 (Rel. 40; Created)  
 DT 15-OCT-2001 (Rel. 40; Last sequence update)  
 DT 15-JUN-2002 (Rel. 41; Last annotation update)  
 DE Extracellular calcium-sensing receptor precursor (Casr) (Parathyroid  
 DE Cell calcium-sensing receptor).  
 GN CASR OR GPRC2A.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID:10090;  
 [1]  
 RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND TISSUE SPECIFICITY.  
 RC STRAIN-C57BL/6; TISSUE-Kidney;  
 RX MEDLINE-20092890; PubMed-10625662;  
 RA Oda Y., Tu C.-L., Chang W., Crumrine D., Koenigues L., Mauro T.,  
 RA Elias P.M., Blike D.D.;  
 RT "The calcium sensing receptor and its alternatively spliced form in  
 RT murine epidermal differentiation";  
 RL J. Biol. Chem. 275:1183-1190(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM A), AND TISSUE SPECIFICITY.  
 RC STRAIN-Black Swiss X 129/SVJ; TISSUE-Kidney;  
 RX MEDLINE-20119279; PubMed-10652312;  
 RA P.M., Garner S.C., Flannery P., Spurney R.F., Quarles L.D.;  
 RT "Sensing of extracellular cations in Casr-deficient osteoblasts.  
 RT Evidence for a novel cation-sensing mechanism";  
 RL J. Biol. Chem. 275:3256-3263(2000).  
 RN [3]  
 RP SEQUENCE OF 256-600 FROM N.A. (ISOFORM A), AND TISSUE SPECIFICITY.  
 RC TISSUE-Epiphyseal cartilage;  
 RX MEDLINE-20043955; PubMed-10579354;  
 RA Chang W., Tu C., Chen T.-H., Komuves L., Oda Y., Pratt S.A.,  
 RA Miller S., Shoback D.;  
 RT "Expression and signal transduction of calcium-sensing receptors in  
 RT cartilage and bone";  
 RL Endocrinology 140:5883-5893(1999).  
 RN [4]  
 RP SEQUENCE OF 507-582 FROM N.A. (ISOFORM A).  
 RC STRAIN-NMRI; TISSUE-Brain;  
 RA Hildengrand J., Ammon H.P.T., Wahl M.A.;  
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE OF 562-814 FROM N.A.  
 RC TISSUE-Kidney;  
 RA Moawad T.I., Riccardi D.;  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE OF 646-799 FROM N.A., AND TISSUE SPECIFICITY.  
 RX MEDLINE-97231187; PubMed-9076582;  
 RA Quarles L.D., Hartle J.E. II, Siddhanti S.R., Guo R., Hinson T.K.;  
 RT "A distinct cation-sensing mechanism in MC3T3-E1 osteoblasts  
 RT functionally related to the calcium receptor";  
 RL J. Bone Miner. Res. 12:393-402(1997).  
 CC -1- FUNCTION: SENSES CHANGES IN THE EXTRACELLULAR CONCENTRATION OF  
 CC CALCIUM IONS. THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY A  
 CC G-PROTEIN THAT ACTIVATES A PHOSPHATIDYLINOSITOL-CALCIUM SECOND  
 CC MESSENGER SYSTEM.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A (SHOWN HERE) AND B; ARE  
 CC PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: EPIDERMIS, KIDNEY AND CARTILAGE.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----

DR EMBL: AF110178; AAD28371.1; -  
 DR EMBL: AF110179; AAD28372.1; -  
 DR EMBL: AF128842; AAD40638.1; -  
 DR EMBL: AF068900; AAC19388.1; -  
 DR EMBL: AB027140; BAA77688.1; -  
 DR EMBL: AF002015; AAC53252.1; -  
 DR EMBL: AF159565; AAF00193.1; -  
 DR MGI:1351351; Gprc2a.  
 DR InterPro: IPR001828; ANF\_receptor.  
 DR InterPro: IPR000337; GPCR\_Mgr.  
 DR Pfam: PF00003; 7tm\_3; 1.  
 DR PRINTS: PR00248; GPCRMR.  
 DR PROSITE: PS00979; G\_PROTEIN\_REC\_P3\_1; 1.  
 DR PROSITE: PS00980; G\_PROTEIN\_REC\_P3\_2; 1.  
 DR PROSITE: PS00981; G\_PROTEIN\_REC\_P3\_3; 1.  
 DR PROSITE: PS50259; G\_PROTEIN\_REC\_P3\_4; 1.  
 KW G-protein coupled receptor; Transmembrane; glycoprotein; Signal;  
 KW Alternative splicing.  
 FT SIGNAL 1 19  
 FT CHAIN 20 1079  
 FT DOMAIN 20 612  
 FT TRANSMEM 613 635  
 FT DOMAIN 636 649  
 FT TRANSMEM 650 670  
 FT DOMAIN 671 681  
 FT TRANSMEM 682 700  
 FT DOMAIN 701 724  
 FT TRANSMEM 725 745  
 FT DOMAIN 746 769  
 FT TRANSMEM 770 792  
 FT DOMAIN 793 805  
 FT TRANSMEM 806 828  
 FT TRANSMEM 829 836  
 FT DOMAIN 837 862  
 FT TRANSMEM 863 1079  
 FT CARBOHYD 90 90  
 FT CARBOHYD 130 130  
 FT CARBOHYD 261 261  
 FT CARBOHYD 287 287  
 FT CARBOHYD 386 386  
 FT CARBOHYD 446 446  
 FT CARBOHYD 468 468  
 FT CARBOHYD 488 488  
 FT CARBOHYD 541 541  
 FT CARBOHYD 594 594  
 FT VARSPLIC 461 537  
 FT CONFLICT 45 45  
 FT CONFLICT 304 304  
 FT CONFLICT 410 410  
 FT CONFLICT 566 566  
 FT CONFLICT 595 595  
 FT CONFLICT 610 610  
 FT CONFLICT 814 814  
 FT CONFLICT 889 889  
 FT CONFLICT 906 909  
 FT CONFLICT 1057 1057  
 FT CONFLICT 1064 1064  
 FT CONFLICT 1076 1076  
 SQ SEQUENCE 1079 AA; 120839 MW; AAF8D8DA72736D6E CRC64;  
 Query Match 65.4%; Score 34; DB 1; Length 1079;  
 Best Local Similarity 62.5%; Pred. No. 62;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 OY 2 ETINCHFK 9  
 DB 354 EFINCHLQ 361  
 RESULT 15  
 CASR\_RAT  
 ID CASR\_RAT STANDARD: PRT; 1079 AA.

AC P48442;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Extracellular calcium-sensing receptor precursor (CaSR) (Parathyroid  
 DE Cell calcium-sensing receptor).  
 GN CaSR OR GPRC2A OR PCARL.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Kidney outer medulla;  
 RX MEDLINE=95116508; PubMed=7816802;  
 RA Riccardi D., Park J., Lee W., Gamba G., Brown E.M., Hebert S.C.;  
 RT "Cloning and functional expression of a rat kidney extracellular  
 RT calcium/polyvalent cation-sensing receptor.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:131-135(1995).  
 RN [2]  
 RP SEQUENCE OF 1-294 FROM N.A.  
 RC STRAIN=Wistar;  
 RX MEDLINE=95241465; PubMed=7724534;  
 RA Ruat M., Snowman A.M., Snyder S.H.;  
 RT "Calcium sensing receptor: molecular cloning in rat and localization  
 RT to nerve terminals.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:3161-3165(1995).  
 CC -1- FUNCTION: SENSE CHANGES IN THE EXTRACELLULAR CONCENTRATION OF  
 CC CALCIUM IONS. THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY A  
 CC G-PROTEIN THAT ACTIVATES A PHOSPHATIDYLINOSITOL-CALCIUM SECOND  
 CC MESSENGER SYSTEM.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC  
 DR EMBL; U10354; AAC52149.1; -;  
 DR EMBL; U20289; AAC52195.1; -;  
 DR InterPro: IPR001828; ANF\_receptor.  
 DR InterPro: IPR000337; GPCR\_Mgr.  
 DR Pfam: PF00003; 7tm\_3; 1.  
 DR Pfam: PF01094; ANF\_receptor; 1.  
 DR PRINTS: PR00248; GPCR\_MGR.  
 DR PROSITE: PS00979; G\_PROTEIN\_REC\_F3\_1; 1.  
 DR PROSITE: PS00980; G\_PROTEIN\_REC\_F3\_2; 1.  
 DR PROSITE: PS00981; G\_PROTEIN\_REC\_F3\_3; 1.  
 DR PROSITE: PS50259; G\_PROTEIN\_REC\_F3\_4; 1.  
 KM G-protein coupled receptor; Transmembrane; Glycoprotein; Signal.  
 FT SIGNAL 1  
 FT CHAIN 20  
 FT DOMAIN 20  
 FT TRANSSEM 613  
 FT DOMAIN 635  
 FT TRANSSEM 636  
 FT TRANSSEM 649  
 FT TRANSSEM 650  
 FT TRANSSEM 671  
 FT TRANSSEM 681  
 FT TRANSSEM 682  
 FT TRANSSEM 700  
 FT TRANSSEM 701  
 FT TRANSSEM 725  
 FT TRANSSEM 746  
 FT TRANSSEM 769  
 FT TRANSSEM 770  
 FT TRANSSEM 792  
 FT TRANSSEM 805  
 FT TRANSSEM 806  
 FT TRANSSEM 828  
 FT TRANSSEM 829  
 FT TRANSSEM 836  
 FT TRANSSEM 837  
 FT TRANSSEM 862  
 FT TRANSSEM 863  
 FT TRANSSEM 1079  
 FT TRANSSEM 90  
 FT TRANSSEM 90  
 FT TRANSSEM 130  
 FT TRANSSEM 130

Query Match	Best Local Similarity	Score 34; DB 1; Length 1079;
Matches 5; Conservative	62.58; 1; Mismatches	2; Indels 0; Gaps 0;
Oy 2 ETINCHK 9	11 111 : 361	
Db 354 ETINCHK 361		

Search completed: July 11, 2003, 11:55:10  
 Job time : 2.86071 secs





```

AC 092035;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Megsln.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21368006; PubMed=11473647;
RA Naganu M., Miyata T., Suzuki D., Umezono T., Hashimoto T., Wada T.,
RA Yagi M., Naganu N., Inagi R., Kuokawa K.;
RT "Cloning of rodent megsln revealed its up-regulation in
RT mesangiolipoliferative nephritis.";
RL Kidney Int. 60:641-652(2001).
CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.
DR EMBL; AF105329; AAL16769.1; -.
DR InterPro; IPR000215; Serpin.
DR Pfam; PF00079; serpin.1.
DR PROSITE; PS00284; SERPIN; UNKNOWN_1.
KM Serpin.
SQ SEQUENCE 380 AA; 42821 MM; D8076CABEE2C2FC64;
Query Match 75.0%; Score 39; DB 11; Length 380;
Best Local Similarity 55.6%; Pred. No. 9.1;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
OY 1 SERPINHX 9
Db 188 SDSLCHFR 196
RESULT 3
O9VDD4 PRELIMINARY; PRT; 199 AA.
ID O9VDD4;
AC O9VDD4;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE CG6300 protein.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abell J.F., Agbayani A., An H.-U., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolintsov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodet A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

```

```

RA Laeko P., Iel Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclob J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003727; AAF55696.1; -.
DR HSSP; P08659; ILIC1.
DR FLYbase; FBgn0038730; CG6300.
DR InterPro; IPR001589; Actbind.actnin.
DR InterPro; IPR000873; AMP-bind.
DR Pfam; PF00501; AMP-binding.1.
DR PROSITE; PS00019; ACTININ_1; UNKNOWN_1.
SQ SEQUENCE 199 AA; 22504 MM; DD6AFEC067AFD5DA CRC64;
Query Match 73.1%; Score 38; DB 5; Length 199;
Best Local Similarity 100.0%; Pred. No. 8.1;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 3 TINCHF 8
Db 3 TINCHF 8
RESULT 4
O9VDD3 PRELIMINARY; PRT; 208 AA.
ID O9VDD3;
AC O9VDD3;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE CG11659 protein.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abell J.F., Agbayani A., An H.-U., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolintsov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodet A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

```

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laško P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mallet B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Metulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,  
 RA Palazzolo M., Plittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Relbert K., Remington K., Saunders R.D.C., Scheller F., Shen H.,  
 RA Shue B.C., Sideri-Klimos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weisslock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003727; AAF55697.1; -  
 DR HSSP: P08659; ILC1.  
 DR Flybase: FBgn0038731; CG11659.  
 DR InterPro: IPR001589; AcbInd\_ActnIn.  
 DR InterPro: IPR000873; AMP-bind.  
 DR Pfam: PF00501; AMP-binding; 1.  
 DR PROSITE: PS00019; ACTININ\_1; UNKNOWN\_1.  
 SQ SEQUENCE 208 AA; 23559 MW; E41495344B417C35 CRC64;

Query Match 73.1%; Score 38; DB 5; Length 208;  
 Best Local Similarity 100.0%; Pred. No. 8.4;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 TINCH 8  
 |||||  
 DB 3 TINCH 8

RESULT 5  
 092J55 PRELIMINARY; PRT; 224 AA.  
 AC 092J55;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Hypothetical protein R03183.  
 GN R03183 OR SMC03800.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Sinorhizobium.  
 OX NCBI\_TaxID=382;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1021;  
 RA MEDLINE=21396507; PubMed=11481430;  
 RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,  
 RA Bolstad P., Becker A., Boutry M., Cadieu E., Drenth S., Gloux S.,  
 RA Godle T., Goffeau A., Kahn D., Kiss E., Lelaune V., Masny D.,  
 RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramseger U.,  
 RA Renard C., Thebaud P., Vandenbol M., Weidner S., Gallbert F.;  
 RT "Analysis of the chromosome sequence of the legume symbiont  
 RT *Sinorhizobium meliloti* strain 1021.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).  
 DR EMBL: AL591793; CAC47762.1; -  
 KW Hypothetical protein; Complete proteome.  
 SO SEQUENCE 224 AA; 24768 MW; 70D206Df4D1AD018 CRC64;

Query Match 71.2%; Score 37; DB 16; Length 224;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 ETINCH 7  
 |||||  
 DB 186 ETINCH 191

RESULT 6  
 09JIP9 PRELIMINARY; PRT; 265 AA.  
 AC 09JIP9;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE Lectin-related NK cell receptor LY49L4 (Natural killer cell receptor  
 DE LY-49L4).  
 GN KLR1A2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CBA/J;  
 RA Makrigianis A.P., Etzler J., Winkler-Pickett R., Ortaido J.R.,  
 RA Anderson S.K.;  
 RT "Identification of the LY49L protein: evidence for activating  
 RT counterparts to inhibitory receptors.";  
 RL J. Leukoc. Biol. 0:0-0(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BA1B/C;  
 RA Silver E.T., Gong D., Hazes B., Kane K.P.;  
 RT "Ly-49W: a novel activating receptor of NOD mice with homology to the  
 RT inhibitory receptor Ly-49G that recognizes H-2Dk and H-2Dd.";  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF204268; AAF75837.1; -  
 DR EMBL: AF307948; AAG41969.1; -  
 DR MGD: MGI:1321091; Klr1a2.  
 DR InterPro: IPR001304; Lectin\_C.  
 DR Pfam: SM00034; CLECT; 1.  
 DR SMART: SM00034; CLECT; 1.  
 DR PROSITE: PS00041; C\_TYPE\_LLECTIN\_2; 1.  
 KW Receptor.  
 SQ SEQUENCE 265 AA; 30497 MW; 0F4B927BF1CFEB8 CRC64;

Query Match 71.2%; Score 37; DB 11; Length 265;  
 Best Local Similarity 75.0%; Pred. No. 16;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 ETINCH 9  
 |||||  
 DB 77 ETINCH 84

RESULT 7  
 09EPB2 PRELIMINARY; PRT; 265 AA.  
 AC 09EPB2;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE Natural killer cell receptor LY49W2 (Natural killer cell receptor Ly-  
 DE 49W2).  
 GN KLR23 OR KLR17 OR LY49W OR LY-49W.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NOD;  
 RA Silver E.T., Gong D.-E., Hazes B., Kane K.P.;  
 RT "Ly49W: a novel activating receptor of NOD mice with homology to the  
 RT inhibitory receptor Ly49G that recognizes H-2Dk and H-2Dd.";  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NOR;

```

RA Silver E.T., Gong D.-E., Hazes B., Kane K.P.;
RT "Ly-49w: a novel activating receptor of NOD mice with homology to the
RT inhibitory receptor Ly-49g that recognizes H-2Dk and H-2Dd.";
RT Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF283251; AAC10158.1; -.
DR EMBL: AF074463; AAC09059.1; -.
DR MGD: MGI:1930932; KlrA23.
DR InterPro: IPR001304; Lectin_C.
DR Pfam: PF00059; Lectin_C.1.
DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS50041; C_TYPE_LECTIN_2; 1.
KW Receptor.
SQ SEQUENCE 265 AA; 30443 MW; 381C38DC9C9A9073 CRC64;

Query Match
Best Local Similarity 71.2%; Score 37; DB 11; Length 265;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 2 ETINCHFK 9
Db 77 ETLNCHDK 84

RESULT 8
O99P19 PRELIMINARY; PRT; 265 AA.
AC O99P19;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Natural killer cell receptor LY49M.
CN KLRA13 OR LY49M.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NOD;
RA MEDLINE=21103262; PubMed=11160290;
RA Silver E.T., Gong D.-E., Hazes B., Kane K.P.;
RT "Ly-49w, an Activating Receptor of Nonobese Diabetic Mice with Close
RT Homology to the Inhibitory Receptor Ly-49g, Recognizes H-2Dk and H-
RT 2Dd.";
RL J. Immunol. 166:2333-2341(2001).
DR EMBL: AF283252; AAk11559.1; -.
DR MGD: MGI:1321090; KlrA13.
DR InterPro: IPR001304; Lectin_C.
DR Pfam: PF00059; Lectin_C.1.
DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS50041; C_TYPE_LECTIN_2; 1.
KW Receptor.
SQ SEQUENCE 265 AA; 30733 MW; F93E1D49EC2670CF CRC64;

Query Match
Best Local Similarity 71.2%; Score 37; DB 11; Length 265;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 2 ETINCHFK 9
Db 77 ETLNCHDK 84

RESULT 9
O9J701 PRELIMINARY; PRT; 268 AA.
AC O9J701;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE Lectin-related NK cell receptor LY49L2 (Natural killer cell receptor
DE Ly-49L2).
CN KLRA12.

```

```

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBA/J;
RA Makrigianis A.P., Etzler J., Winkler-Pickett R., Ortaldo J.R.,
RA Anderson S.K.;
RT "Identification of the Ly49L protein: evidence for activating
RT counterparts to inhibitory receptors.";
RL J. Leukoc. Biol. 0:0-0(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RA Silver E.T., Gong D., Hazes B., Kane K.P.;
RT "Ly-49w: a novel activating receptor of NOD mice with homology to the
RT inhibitory receptor Ly-49g that recognizes H-2Dk and H-2Dd.";
RT Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF204266; AAF75835.1; -.
DR EMBL: AF307947; AAC41968.1; -.
DR MGD: MGI:1321091; KlrA12.
DR InterPro: IPR001304; Lectin_C.
DR Pfam: PF00059; Lectin_C.1.
DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS50041; C_TYPE_LECTIN_2; 1.
KW Receptor.
SQ SEQUENCE 268 AA; 30786 MW; D15013DF7417D4AA CRC64;

Query Match
Best Local Similarity 71.2%; Score 37; DB 11; Length 268;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 2 ETINCHFK 9
Db 80 ETLNCHDK 87

RESULT 10
O9EP57 PRELIMINARY; PRT; 268 AA.
AC O9EP57;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Natural killer cell receptor LY49W1 (Natural killer cell receptor Ly-
DE 49W1).
CN KLRA23 OR KLRA17 OR LY49W OR LY-49W.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NOD;
RA Silver E.T., Gong D.-E., Hazes B., Kane K.P.;
RT "Ly49w: a novel activating receptor of NOD mice with homology to the
RT inhibitory receptor Ly49g that recognizes H-2Dk and H-2Dd.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NOR;
RA Silver E.T., Gong D.-E., Hazes B., Kane K.P.;
RT "Ly-49w: a novel activating receptor of NOD mice with homology to the
RT inhibitory receptor Ly-49g that recognizes H-2Dk and H-2Dd.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF283250; AAC10157.1; -.
DR EMBL: AF074459; AAC09058.1; -.
DR MGD: MGI:1930932; KlrA23.
DR InterPro: IPR001304; Lectin_C.
DR Pfam: PF00059; Lectin_C.1.
DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS50041; C_TYPE_LECTIN_2; 1.

```

KW Receptor.  
SQ SEQUENCE 268 AA; 30733 MW; E607B9781942AA61 CRC64;

Query Match 71.2%; Score 37; DB 11; Length 268;  
Best Local Similarity 75.0%; Pred. No. 17;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 ETINCHFK 9  
11:11111  
DB 80 ETLNCHDK 87

## RESULT 11

O9J100

ID 09J100 PRELIMINARY; PRT; 278 AA.

AC 09J100; 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE Lectin-related NK cell receptor LY49L3.

GN KLRA12.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CBA/J;

RA Makrigianis A.P., Etzler J., Winkler-Pickett R., Ortaldo J.R.,

RA Anderson S.K.;

RT \*Identification of the LY49L protein: evidence for activating

RT counterparts to inhibitory receptors.;

RL J. Leukoc Biol. 0:0-0(2000).

DR EMBL: AF204267; AAF75836.1; -.

DR MGD: MGI:1321091; Klra12.

DR InterPro: IPR001304; Lectin\_C.

DR Pfam: PF000059; lectin\_c; 1.

DR SMART: SM00034; CLECT; 1.

DR PROSITE: PS50041; C\_Type\_Lectin\_2; 1.

KW Receptor.

SQ SEQUENCE 278 AA; 31885 MW; 5F1B09E0FD89EA80 CRC64;

Query Match 71.2%; Score 37; DB 11; Length 278;

Best Local Similarity 75.0%; Pred. No. 17;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 ETINCHFK 9  
11:11111  
DB 77 ETLNCHDK 84

## RESULT 12

O9J102

ID 09J102 PRELIMINARY; PRT; 281 AA.

AC 09J102; 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE Lectin-related NK cell receptor LY49L1.

GN KLRA12.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CBA/J;

RA Makrigianis A.P., Etzler J., Winkler-Pickett R., Ortaldo J.R.,

RA Anderson S.K.;

RT \*Identification of the LY49L protein: evidence for activating

RT counterparts to inhibitory receptors.;

RL J. Leukoc Biol. 0:0-0(2000).

DR EMBL: AF204265; AAF75834.1; -.

DR MGD: MGI:1321091; Klra12.  
DR InterPro: IPR001304; Lectin\_C.

DR Pfam: PF000059; lectin\_c; 1.

DR SMART: SM00034; CLECT; 1.

DR PROSITE: PS50041; C\_Type\_Lectin\_2; 1.

KW Receptor.

SQ SEQUENCE 281 AA; 32175 MW; 2589356CB9F4DE09 CRC64;

Query Match 71.2%; Score 37; DB 11; Length 281;  
Best Local Similarity 75.0%; Pred. No. 17;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 ETINCHFK 9  
11:11111  
DB 80 ETLNCHDK 87

## RESULT 13

O12180

ID 012180 PRELIMINARY; PRT; 1030 AA.

AC 012180; 01-NOV-1996 (TREMBlrel. 01, Created)

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)

DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE Chromosome XV reading frame ORF YOL089C.

GN HAL9 OR YOL089C.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.

OX NCBI\_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A.

RA Zumbstein E., Pearson B.M., Kalogeropoulos A., Schweizer M.;

RA Submitted (JUL-1996) to the EMBL/Genbank/DBJ databases.

RL [2]

RP SEQUENCE FROM N.A.

RA MIPS: Submitted (JUL-1996) to the EMBL/Genbank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=FY1679;

RX MEDLINE=96021609; PubMed=8533473;

RA Zumbstein E., Pearson B.M., Kalogeropoulos A., Schweizer M.;

RT "A 29.425 kb segment on the left arm of yeast chromosome XV contains

RT more than twice as many unknown as known open reading frames.;"

CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).

CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE BINUCLEAR CLUSTER DOMAIN.

DR EMBL: Z74831; CAA99101.1; -.

DR EMBL: X83121; CAA58190.1; -.

DR HSSP: P21228; 2AUC.

DR SGD: S0005449; HAL9.

DR InterPro: IPR001138; Fungi\_TRN.

DR Pfam: PF00172; zn\_c1us; 1.

DR SMART: SM00066; GAL4; 1.

DR PROSITE: PS00463; ZN2\_CV6\_FUNGAL\_1; 1.

DR PROSITE: PS50048; ZN2\_CV6\_FUNGAL\_2; 1.

KW DNA-binding; Metal-binding; Nuclear protein; Transcription regulation;

KW Zinc.

SQ SEQUENCE 1030 AA; 117925 MW; 8C8BDE8406D2105 CRC64;

Query Match 71.2%; Score 37; DB 3; Length 1030;  
Best Local Similarity 55.6%; Pred. No. 54;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 SETINCHFK 9  
1:11111  
DB 689 SRVLNCHR 697

## RESULT 14

O9V9N7

ID 09V9N7 PRELIMINARY; PRT; 1354 AA.

AC Q9Y9N7; 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)  
 DE CG6448 protein.  
 GN CG6448.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 RX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazet J.R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Bens P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Botkova D., Botchan D.R., Bouck J., Brokstein P., Brotlier P.,  
 RA Butlis K.C., Busam D.A., Butler H., Cadieu L.B., Davies P.,  
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferrar C., Ferrier S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jajani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacle D.J.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter A., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003782; AAF57249.1; -  
 DR FLYBase: FBgn0032976; CG6448.  
 SQ SEQUENCE 1354 AA; 150276 MW; 770F5904A634EFD6 CRC64;

Query Match 69.2%; Score 36; DB 5; Length 1354;  
 Best Local Similarity 83.3%; Pred. No. 1.1e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 ETINCH 7  
 11:111  
 Db 1172 ETVNCH 1177

RESULT 15  
 ID Q856H9 PRELIMINARY; PRT: 1739 AA.  
 AC Q856H9;  
 DT 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE Putative retroelement.  
 GN OSJNB0023M1.15.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 RX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NTIPPOBARE;  
 RA McCombie W.R., de la Bastide M., Spiegel L., Preston R., Kirchoff K.,  
 RA Kuit K., Nascimben L., Baker J., Santos L., Zutavern T., Miller B.,  
 RA Cunniff D.M., Katzenberger F., Muller S., Bell M., Ballija V., Shah R.,  
 RA King L., Yang C., Dike S., O'Shaughnessy A., Palmer L., Dedhia N.;  
 RT "Genomic sequence for Oryza sativa, Nipponbare strain, clone  
 OSJNB0023M1, from chromosome 10, complete sequence";  
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC092749; AAM08361.1; -  
 SQ SEQUENCE 1739 AA; 196006 MW; 47F7707D33152286 CRC64;

Query Match 69.2%; Score 36; DB 10; Length 1739;  
 Best Local Similarity 50.0%; Pred. No. 1.3e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 2 ETINCH 9  
 11:11111  
 Db 1035 QSVNCHFO 1042

Search completed: July 11, 2003, 11:58:48  
 Job time : 5.92931 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:42:16 ; Search time 4.01663 Seconds  
(without alignments)  
232.223 Million cell updates/sec

Title: US-10-091-442-2  
Perfect score: 40  
Sequence: 1 WQSAFTK 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A.Geneseq\_101002:\*

1:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
2:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
3:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
4:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
5:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
6:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
7:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
8:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
9:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
10:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
11:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
12:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
13:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
14:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
15:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
16:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
17:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
18:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
19:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
20:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
21:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
22:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*
23:	/SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	100.0	380	15	AA198379 Human megakaryocyte
2	40	100.0	380	20	AA198254 Human megaslin prote
3	40	100.0	380	21	AA198241 Human megaslin prote
4	40	100.0	380	22	AA1983075 Human megaslin prote
5	36	90.0	306	22	ABG02777 Novel human diagno
6	36	90.0	323	22	ABG11497 Novel human diagno
7	36	90.0	368	20	AA198256 Mouse megaslin prote
8	36	90.0	368	21	AA198241 Mouse megaslin prote
9	36	90.0	368	22	AA1983077 Murine megaslin prote
10	36	90.0	380	20	AA198255 Rat megaslin protein

11	36	90.0	380	21	AA19824150
12	36	90.0	380	22	AA1984286
13	36	90.0	380	22	AA1983076
14	32	80.0	219	22	ABG29897
15	32	80.0	331	22	ABG18342
16	32	80.0	931	21	AA1982833
17	32	80.0	938	20	AA1980528
18	32	80.0	940	20	AA1983082
19	32	80.0	944	21	AA1984351
20	32	80.0	959	22	AA1984950
21	32	80.0	1072	22	AA1986216
22	32	77.5	46	21	AA19828753
23	31	77.5	119	22	AA19806492
24	31	77.5	121	21	AA19858546
25	31	77.5	125	22	AA19891461
26	31	77.5	168	22	AB110348
27	31	77.5	187	21	AA19837074
28	31	77.5	187	21	AA19852578
29	31	77.5	190	21	AA19852577
30	31	77.5	191	21	AA19837073
31	31	77.5	193	20	AA19838921
32	31	77.5	194	22	AA198012
33	31	77.5	202	21	AA19806264
34	31	77.5	264	20	AA19835068
35	31	77.5	356	20	AA19838925
36	31	77.5	357	20	AA19838922
37	31	77.5	357	20	AA19838923
38	31	77.5	364	20	AA19838924
39	31	77.5	514	22	AA19896264
40	31	77.5	666	23	AA19829064
41	31	77.5	733	20	AA19835373
42	31	77.5	813	21	AA1982835
43	31	77.5	978	21	AA1982834
44	31	77.5	978	22	AA1983886
45	31	77.5	978	23	AA1980543

## ALIGNMENTS

RESULT 1	
AA198379	standard; Protein; 380 AA.
ID	AA198379
AC	AA198379;
XX	
DT	16-AUG-1994 (first entry)
XX	
DE	Human megakaryocyte differentiation factor.
XX	
KW	Human megakaryocyte differentiation factor; MDF; thrombopoietin;
KW	hematopoietic stimulating factor; thrombocytopoietin; platelet;
XX	bone marrow transplantation; cancer chemotherapy.
OS	Homo sapiens.
PN	EP583884-A.
PD	23-FEB-1994.
XX	
PF	19-JUL-1993; 93EP-0305654.
XX	
PR	17-JUL-1992; 92JP-0212305.
XX	
PA	04-MAR-1993; 93JP-0067339.
XX	
PI	(SUNR) SUNTORY LTD.
XX	(TSUJ) TSUJIMOTO M.
PI	Ishida N, Iwasa F, Kurihara T, Miura K, Nakazato H;
XX	TSUJIMOTO M, TSURUOKA N, Yamaguchi N, Yamachi K;
DR	WPI; 1994-058782/08.
DR	N-PSDB; AA056670.

Rat megaslin protein  
Rat megaslin protein  
Rat megaslin protein  
Novel human diagno  
Novel human diagno  
C. pneumoniae CPN1  
Chlamydia pneumoniae  
Human ORF3115  
Human ORF3115  
Human ORF3115  
Drosophila melanog  
Protein fragment e  
Human polypeptide  
Arabidopsis thaliana  
C. glutamicum prote  
Human cDNA SEQ ID  
Arabidopsis thaliana  
Arabidopsis thaliana  
Arabidopsis thaliana  
Arabidopsis thaliana  
Neisseria meningit  
Human immunoglobul  
Arabidopsis thaliana  
Chlamydia pneumoniae  
Neisseria gonorrhoe  
Neisseria gonorrhoe  
Neisseria meningit  
Neisseria meningit  
Putative P. abyssal  
Streptococcus poly  
Amino acid sequenc  
C. pneumoniae CPN1  
C. pneumoniae CPN1  
C. pneumoniae CPN1  
Chlamydia pneumoniae

```

XX New megakaryocyte differentiation factor - isolated from human
PT epidermoid carcinoma cells, used to treat conditions involving a
XX decrease in platelets
PS Claim 7; Page 30-32; 47pp; English.
XX
CC Human MDF can be isolated from a culture of human epidermoid
CC carcinoma A431 cells in protein-free medium. The MDF stimulates
CC differentiation of megakaryocytes from myeloid cells in the presence
CC of IL-3. The MDF acts in vivo as a thrombopoietin making it useful
CC for treatment of diseases involving a decrease in platelet number
CC (esp. thrombocytopaenia) such as occurs in bone marrow
CC transplantation and in chemotherapy. MDF has mol.wt. 55-57kd by
CC SDS-PAGE and exhibits an isoelectric point of 6.0-7.0.
XX
SQ Sequence 380 AA;
Query Match 100.0%; Score 40; DB 15; Length 380;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WQSAFTK 7
DB 181 WQSAFTK 187

RESULT 2
AA08254
ID AA08254 standard; Protein: 380 AA.
XX
AC AAY08254;
XX
DT 14-JUL-1999 (first entry)
XX
DE Human megsin protein.
XX
KW Megsin; mesangial cell; treatment; diagnosis; disease; Iga nephropathy;
XX human; rat; murine.
XX
OS Homo sapiens.
XX
PN W09915652-A1.
XX
PD 01-APR-1999.
XX
PF 22-SEP-1998; 98WO-JP04269.
XX
PR 22-SEP-1997; 97JP-0275302.
XX
PA (KURO/) KUROKAWA K.
XX (MIYA/) MIYATA T.
XX
PI Miyata T;
XX
DR WPI: 1999-276983/23.
XX N-PSDB; AAX56712.
XX
PT Megsin protein expressed specifically in mesangial cells
XX
PS Claim 1; Page 62-64; 100pp; Japanese.
XX
CC This invention describes the isolation of novel megsin nucleic acid and
CC proteins from human, rat and mouse tissue. This protein is expressed
CC specifically in mesangial cells. The products of the invention are
CC useful for the treatment and diagnosis of diseases involving mesangial
CC cells, such as Iga nephropathy.
XX
SQ Sequence 380 AA;
Query Match 100.0%; Score 40; DB 20; Length 380;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 WQSAFTK 7
DB 181 WQSAFTK 187

RESULT 3
AAB24142
ID AAB24142 standard; Protein: 380 AA.
XX
AC AAB24142;
XX
DT 30-JAN-2001 (first entry)
XX
DE Human megsin protein sequence SEQ ID NO:2.
XX
KW Megsin; mesangium-predominant gene; serpin regulated; nephropathy;
KW Iga; immunoglobulin A; detection; renal function; renal disorder;
XX diagnosis; biological sample; blood; urine.
XX
OS Homo sapiens.
XX
PN W0200057189-A1.
XX
PD 28-SEP-2000.
XX
PF 17-MAR-2000; 2000WO-JP01646.
XX
PR 19-MAR-1999; 99JP-0075305.
XX 28-OCT-1999; 99JP-0306623.
XX
PA (KURO/) KUROKAWA K.
XX (FUSO ) FUSO PHARM IND LTD.
XX (MIYA/) MIYATA T.
XX
PI Miyata T;
XX
DR WPI: 2000-611642/58.
XX N-PSDB; AAA99294.
XX
PT Evaluating renal function comprises assaying megsin protein in
XX biological sample
XX
PS Example 2; Page 66-69; 93pp; Japanese.
XX
CC The present invention describes a method for evaluating renal function.
CC The method comprises assaying megsin protein in biological sample. Also
CC described are: (1) use of an anti-megsin protein antibody for diagnosing
CC renal function; and (2) a kit for detecting megsin protein comprising:
CC (a) anti-megsin protein antibody attached to solid magnetic particles;
CC (b) direct or indirect fixing for the antibody to the particles; and
CC (c) a magnet. The process is useful for evaluating renal function and
CC diagnosing renal disorders by assaying megsin protein in biological
CC samples (preferably urine or blood). The process is reproducible and
CC gives accurate results. The present sequence represents the human megsin
CC protein, which is given in the exemplification of the present invention.
XX
SQ Sequence 380 AA;
Query Match 100.0%; Score 40; DB 21; Length 380;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WQSAFTK 7
DB 181 WQSAFTK 187

RESULT 4
AAB83075
ID AAB83075 standard; Protein: 380 AA.
XX
AC AAB83075;

```



XX 10-JUL-2001 (first entry)  
XX  
XX  
DE Human mesgln protein.  
XX  
KW Human: mesgln; mesangial cell proliferative nephritis; nephrotropic;  
KW transgenic mouse; glomerular disease; animal model; drug screening.  
XX  
OS Homo sapiens.  
XX  
PM WO200124628-A1.  
XX  
PD 12-APR-2001.  
XX  
PF 06-OCT-2000; 2000WO-JP06988.  
XX  
PR 06-OCT-1999; 99JP-0285736.  
XX  
PA (KURO/) KUROKAWA K.  
PA (MIYA/) MIYATA T.  
XX  
PI Miyata T;  
XX  
DR WPI: 2001-300136/31.  
DR N-PSDB: AAF82438.  
XX  
PT Mouse model for mesangial cell proliferative nephritis for development  
PT and screening of new treatments -  
XX  
PS Example 4; Page 44-46; 62pp; Japanese.  
XX  
CC The present sequence is human mesgln. The human mesgln coding  
CC sequence may be introduced into a mouse to produce an animal model of  
CC mesangial cell proliferative nephritis. The symptoms include  
CC enlargement of the mesangial base region, sedimentation of an immune  
CC complex and an increase in mesangial cells. The animal model is useful  
CC for analyzing the pathology of chronic glomerular diseases and for  
CC screening compositions for prevention and treatment of the diseases.  
CC Highly uniform models can be made easily and in large numbers using  
CC this method.  
XX  
SQ Sequence 380 AA:  
XX  
Query Match 100.0%; Score 40; DB 22; Length 380;  
Best Local Similarity 100.0%; Pred. NO. 12;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 WQSAFTK 7  
DB 181 WQSAFTK 187  
XX  
RESULT 5  
ABG02777  
ID ABC02777 standard; Protein: 306 AA.  
XX  
AC ABC02777;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #2768.  
XX  
KW Human: chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PM WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX

PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI: 2001-639362/73.  
DR N-PSDB: AAS66964.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 20; SEQ ID NO 33136; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 306 AA:  
XX  
Query Match 90.0%; Score 36; DB 22; Length 306;  
Best Local Similarity 85.7%; Pred. NO. 54;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 1 WQSAFTK 7  
DB 108 WQSAFTK 114  
XX  
RESULT 6  
ABG11497  
ID ABG11497 standard; Protein: 323 AA.  
XX  
AC ABG11497;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #11488.  
XX  
KW Human: chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PM WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX

XX (HXSE-) HXSEQ INC.  
 PA Drmanac RT, Liu C, Tang YT;  
 PI WPI: 2001-639362/73.  
 XX N-PSDB: AAS75684.  
 DR  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity  
 PS Claim 20: SEQ ID No 41856; 103pp; English.  
 XX  
 PS The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridization probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 SQ Sequence 323 AA;  
 OY  
 Db 108 MOSVFTK 114  
 OY 1 MOSAFTK 7  
 Db 108 MOSVFTK 114  
 RESULT 7  
 AAY08256  
 ID AAY08256 standard; Protein: 368 AA.  
 XX  
 AC AAY08256;  
 XX  
 DT 14-JUL-1999 (first entry)  
 XX  
 DE Mouse megalin protein.  
 XX  
 KM Megalin; mesangial cell; treatment; diagnosis; disease; IGA nephropathy;  
 KW human; rat; murine.  
 XX  
 OS Mus musculus.  
 XX  
 FT Key Location/Qualifiers  
 FT Protein 1..368  
 FT /note= "partial sequence"  
 XX  
 PN WO9915652-A1.  
 PD 01-APR-1999.  
 XX  
 PF 22-SEP-1998; 98WO-JP04269.  
 XX

PR 22-SEP-1997; 97JP-0275302.  
 XX  
 PA (KURO/) KUROKAWA K.  
 PA (MIYA/) MIYATA T.  
 XX  
 PI Miyata T;  
 XX  
 DR WPI: 1999-276983/23.  
 DR N-PSDB: AAX56714.  
 XX  
 PT Megalin protein expressed specifically in mesangial cells  
 PS Claim 1; Page 76-79; 100pp; Japanese.  
 XX  
 PS This invention describes the isolation of novel megalin nucleic acid and  
 CC proteins from human, rat and mouse tissue. This protein is expressed  
 CC specifically in mesangial cells. The products of the invention are  
 CC useful for the treatment and diagnosis of diseases involving mesangial  
 CC cells, such as IGA nephropathy.  
 CC  
 SQ Sequence 368 AA;  
 OY  
 Db 169 WKSFTK 175  
 OY 1 MOSAFTK 7  
 Db 169 WKSFTK 175  
 RESULT 8  
 AAB24151  
 ID AAB24151 standard; Protein: 368 AA.  
 XX  
 AC AAB24151;  
 XX  
 DT 30-JAN-2001 (first entry)  
 XX  
 DE Mouse megalin protein sequence SEQ ID NO:21.  
 XX  
 KM Megalin; mesangium-predominant gene; serpin regulated; nephropathy;  
 KW IGA; immunoglobulin A; detection; renal function; renal disorder;  
 KW diagnosis; biological sample; blood; urine.  
 XX  
 OS Mus musculus.  
 XX  
 PN WO200057189-A1.  
 PN 28-SEP-2000.  
 PD 17-MAR-2000; 2000WO-JP01646.  
 PF 19-MAR-1999; 99JP-0075305.  
 PR 28-OCT-1999; 99JP-0306623.  
 XX  
 PA (KURO/) KUROKAWA K.  
 PA (FUSO) FUSO PHARM IND LTD.  
 PA (MIYA/) MIYATA T.  
 XX  
 PI Miyata T;  
 XX  
 DR WPI: 2000-611642/58.  
 DR N-PSDB: AAC55239.  
 XX  
 FT Evaluating renal function comprises assaying megalin protein in  
 FT biological sample  
 XX  
 PS Disclosure; Page 89-91; 93pp; Japanese.  
 XX  
 CC The present invention describes a method for evaluating renal function.  
 CC The method comprises assaying megalin protein in biological sample. Also  
 CC described are: (1) use of a anti-megalin protein antibody for diagnosing

CC renal function; and (2) a kit for detecting mesgin protein comprising:  
 CC (a) anti-mesgin protein antibody attached to solid magnetic particles;  
 CC (b) direct or indirect fixing for the antibody to the particles; and  
 CC (c) a magnet. The process is useful for evaluating renal function and  
 CC diagnosing renal disorders by assaying mesgin protein in biological  
 CC samples (preferably urine or blood). The process is reproducible and  
 CC gives accurate results. The present sequence represents the mouse mesgin  
 CC protein, which is given in the exemplification of the present invention.  
 XX

SO Sequence 368 AA:

Query Match 90.0%; Score 36; DB 21; Length 368;  
 Best Local Similarity 85.7%; Pred. No. 65;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WQSAFTK 7  
 1:|||||  
 Db 169 WKSFTK 175

RESULT 9  
 AAB83077 ID AAB83077 standard; Protein: 368 AA.  
 AC AAB83077;  
 XX 10-JUL-2001 (first entry)  
 DT  
 XX Murine mesgin protein.  
 DE  
 XX Mouse: mesgin; mesangial cell proliferative nephritis; nephrotropic;  
 KW transgenic mouse; glomerular disease; animal model; drug screening.  
 XX  
 OS Mus musculus.  
 XX  
 XX WO200124628-A1.  
 PN  
 XX 12-APR-2001.  
 PD  
 XX  
 XX 06-OCT-2000; 2000MO-JP06988.  
 PE  
 XX 06-OCT-1999; 99JP-0285736.  
 PR  
 XX (KUROO/) KUROKAWA K.  
 PA (MIYA/) MIYATA T.  
 XX  
 PI Miyata T;  
 PL  
 XX WPI: 2001-300136/31.  
 DR  
 XX N-PSDB: AAF82440.  
 DR  
 XX Mouse model for mesangial cell proliferative nephritis for development  
 PT and screening of new treatments -  
 PT  
 XX  
 PS Disclosure: Page 52-53; 62pp; Japanese.  
 PS  
 XX The present sequence is murine mesgin protein. The human mesgin coding  
 CC sequence may be introduced into a mouse to produce an animal model of  
 CC mesangial cell proliferative nephritis. The symptoms include  
 CC enlargement of the mesangial base region, sedimentation of an immune  
 CC complex and an increase in mesangial cells. The animal model is useful  
 CC for analyzing the pathology of chronic glomerular diseases and for  
 CC screening compositions for prevention and treatment of the diseases.  
 CC Highly uniform models can be made easily and in large numbers using  
 CC this method.  
 CC  
 SQ Sequence 368 AA:

Query Match 90.0%; Score 36; DB 22; Length 368;  
 Best Local Similarity 85.7%; Pred. No. 65;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WQSAFTK 7

Db 169 WKSFTK 175  
 1:|||||

RESULT 10  
 AAY08255 ID AAY08255 standard; Protein: 380 AA.  
 XX  
 AC AAY08255;  
 XX  
 XX 14-JUL-1999 (first entry)  
 DT  
 XX Rat mesgin protein.  
 DE  
 XX Mesgin; mesangial cell; treatment; diagnosis; disease; Iga nephropathy;  
 KW human; rat; murine.  
 KW  
 XX  
 XX Rattus rattus.  
 OS  
 XX  
 XX WO9915652-A1.  
 PN  
 XX 01-APR-1999.  
 PD  
 XX 22-SEP-1998; 98WO-JP04269.  
 PE  
 XX 22-SEP-1997; 97JP-0275302.  
 PR  
 XX (KUROO/) KUROKAWA K.  
 PA (MIYA/) MIYATA T.  
 XX  
 XX Miyata T;  
 PI  
 XX WPI: 1999-276983/23.  
 DR  
 XX N-PSDB: AAX56712.  
 DR  
 XX Mesgin protein expressed specifically in mesangial cells  
 XX  
 XX Claim 1; Page 69-72; 100pp; Japanese.  
 PS  
 XX This invention describes the isolation of novel mesgin nucleic acid and  
 CC proteins from human, rat and mouse tissue. This protein is expressed  
 CC specifically in mesangial cells. The products of the invention are  
 CC useful for the treatment and diagnosis of diseases involving mesangial  
 CC cells, such as Iga nephropathy.  
 CC  
 XX  
 SQ Sequence 380 AA:

Query Match 90.0%; Score 36; DB 20; Length 380;  
 Best Local Similarity 85.7%; Pred. No. 67;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WQSAFTK 7  
 1:|||||  
 Db 181 WKSFTK 187

RESULT 11  
 AAB24150 ID AAB24150 standard; Protein: 380 AA.  
 XX  
 AC AAB24150;  
 XX  
 XX 30-JAN-2001 (first entry)  
 DT  
 XX Rat mesgin protein sequence SEQ ID NO:19.  
 DE  
 XX  
 XX Mesgin; mesangium-predominant gene; serpin regulated; nephropathy;  
 KW Iga; immunoglobulin A; detection; renal function; renal disorder;  
 KW diagnosis; biological sample; blood; urine.  
 OS  
 XX Rattus norvegicus.  
 XX  
 FH Key Location/Qualifiers

FT	Misc-difference	51	/note= "unspecified"
FT	Misc-difference	94	/note= "unspecified"
FT	Misc-difference	94	/note= "unspecified"
XX	WO200057189-A1.		
XX	28-SEP-2000.		
XX	17-MAR-2000; 2000WO-JP01646.		
XX	19-MAR-1999; 99JP-0075305.		
PR	28-OCT-1999; 99JP-0306623.		
XX	(KURO/ KUROKAWA K.		
PA	(FUSO ) FUSO PHARM IND LTD.		
PA	(MIYA/ ) MIYATA T.		
XX	Miyata T;		
PI	WPI: 2000-611642/58.		
DR	N-PSDB; AAC55238.		
XX			
XX	Evaluating renal function comprises assaying megin protein in		
PT	biological sample		
XX			
PS	Example 2; Page 81-84; 93pp; Japanese.		
XX			
CC	The present invention describes a method for evaluating renal function.		
CC	The method comprises assaying megin protein in biological sample. Also		
CC	described are: (1) use of a anti-megin protein antibody for diagnosing		
CC	renal function; and (2) a kit for detecting megin protein comprising:		
CC	(a) anti-megin protein antibody attached to solid magnetic particles;		
CC	(b) direct or indirect fixing for the antibody to the particles; and		
CC	(c) a magnet. The process is useful for evaluating renal function and		
CC	diagnosing renal disorders by assaying megin protein in biological		
CC	samples (preferably urine or blood). The process is reproducible and		
CC	gives accurate results. The present sequence represents the rat megin		
CC	protein, which is given in the exemplification of the present invention.		
XX			
XX	Sequence 380 AA;		
SQ			
OY	1 WOSAFTK 7		
DB	181 WKSAPTK 187		
XX			
XX	RESULT 12		
XX	AAG64286		
XX	AAG64286 standard; Protein; 380 AA.		
XX	AAG64286;		
XX	21-SEP-2001 (first entry)		
XX			
XX	Rat megin protein.		
XX			
XX	Rat; megin; renal mesangial cell; mesangium proliferative nephritis.		
XX	Rattus norvegicus.		
OS	WO200148019-A1.		
XX			
XX	05-JUL-2001.		
XX			
XX	26-DEC-2000; 2000WO-JP09251.		
XX			
XX	28-DEC-1999; 99JP-0373677.		
XX			

PA (KUROO/) KUROKAWA K.  
 PA (MIYA/) MIYATA T.  
 XX  
 XX  
 PI Miyata T;  
 XX  
 DR WPI: 2001-425651/45.  
 DR N-PSDB: AAH48181.  
 XX  
 PT New antibody recognizing a partial sequence of rat mesgin protein for  
 XX diagnosis of mesangium proliferative nephritis -  
 PS  
 XX Disclosure: Page 54-56; 63pp; Japanese.  
 CC  
 CC The present invention relates to a novel antibody which recognises a  
 CC peptide consisting of residues 341-354 of rat mesgin protein. The present  
 CC sequence is the protein sequence for rat mesgin, which was used in the  
 CC present invention. Mesgin is highly expressed in renal mesangial cells  
 CC and its level is elevated in mesangium proliferative nephritis. Assay of  
 CC the serum or urine level using the antibody is therefore indicative of  
 CC this type of disorder.  
 XX  
 SQ Sequence 380 AA;  
 QY  
 Query Match 90.0%; Score 36; DB 22; Length 380;  
 Best Local Similarity 85.7%; Pred No. 67;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0.  
 1 WQSAFTK 7  
 1:|||||  
 181 WKSATFK 187  
 Db  
 RESULT 13  
 AAB83076  
 ID AAB83076 standard; Protein; 380 AA.  
 XX  
 AC AAB83076;  
 XX  
 DT 10-JUL-2001 (first entry)  
 XX  
 DE Rat mesgin protein.  
 XX  
 KW Rat; mesgin; mesangial cell proliferative nephritis; nephrotropic;  
 KW Transgenic mouse; glomerular disease; animal model; drug screening.  
 XX  
 OS Rattus norvegicus.  
 XX  
 PN WO200124628-A1.  
 XX  
 PD 12-APR-2001.  
 XX  
 PF 06-OCT-2000; 2000WO-JP06988.  
 XX  
 PR 06-OCT-1999; 99JP-0285736.  
 XX  
 PA (KUROO/) KUROKAWA K.  
 PA (MIYA/) MIYATA T.  
 XX  
 XX  
 PI Miyata T;  
 XX  
 DR WPI: 2001-300136/31.  
 DR N-PSDB: AAF82439.  
 XX  
 PT Mouse model for mesangial cell proliferative nephritis for development  
 XX and screening of new treatments -  
 PS  
 XX Disclosure: Page 48-50; 62pp; Japanese.  
 CC  
 CC The present sequence is rat mesgin. The human mesgin coding  
 CC sequence may be introduced into a mouse to produce an animal model of  
 CC mesangial cell proliferative nephritis. The symptoms include  
 CC enlargement of the mesangial base region, sedimentation of an immune  
 CC complex and an increase in mesangial cells. The animal model is useful

CC for analysing the pathology of chronic glomerular diseases and for  
 CC screening compositions for prevention and treatment of the diseases.  
 CC Highly uniform models can be made easily and in large numbers using  
 CC this method.

XX Sequence 380 AA;

Query Match 90.0%; Score 36; DB 22; Length 380;  
 Best Local Similarity 85.7%; Pred. No. 67;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WQSAFTK 7  
 1:11111  
 Db 181 WKSAAFTK 187

RESULT 14  
 ABG29897  
 ID ABG29897 standard; Protein; 219 AA.

XX AC ABG29897;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #29888.

XX Human: chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI: 2001-639362/73.

DR N-PSDB: AAS94084.

PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity

PS Claim 20; SEQ ID No 60256; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 219 AA;

Query Match 80.0%; Score 32; DB 22; Length 219;  
 Best Local Similarity 71.4%; Pred. No. 2.2e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 WQSAFTK 7  
 1:11111  
 Db 196 WESEFTK 202

RESULT 15  
 ABG18342  
 ID ABG18342 standard; Protein; 331 AA.

XX AC ABG18342;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #18331.

XX Human: chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI: 2001-639362/73.

DR N-PSDB: AAS82529.

PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity

PS Claim 20; SEQ ID No 48701; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 331 AA;

Query Match 80.0%; Score 32; DB 22; Length 331;  
 Best Local Similarity 71.4%; Pred. No. 3.3e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 WQSAFTK 7  
 - 1:1 111  
 Db 308 WESEFTK 314

Search completed: July 11, 2003, 11:54:18  
 Job time : 5.01663 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:49:36 ; Search time 1.25156 seconds

(without alignments)  
164.563 Million cell updates/sec

Title: US-10-091-442-2

Perfect score: 40

Sequence: 1 WQSAFTK 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*

1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*

2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*

3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*

4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*

5: /cgn2\_6/ptodata/1/1aa/PTCUS.COMB.pep:\*

6: /cgn2\_6/ptodata/1/1aa/Dackfilest1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	100.0	7	2	US-08-611-977-2
2	40	100.0	380	2	US-08-472-659-34
3	40	100.0	380	2	US-08-474-661-34
4	40	100.0	380	2	US-08-611-977-34
5	30	75.0	385	3	US-09-053-866-2
6	30	75.0	385	4	US-09-479-130-2
7	29	72.5	381	4	US-09-216-295-22
8	28	70.0	146	1	US-08-543-238-11
9	28	70.0	146	1	US-08-420-526-11
10	28	70.0	177	4	US-09-134-001C-4482
11	28	70.0	189	2	US-08-483-695-46
12	28	70.0	189	2	US-07-965-285-46
13	28	70.0	189	2	US-08-487-231-46
14	28	70.0	189	4	US-09-201-912-46
15	28	70.0	269	4	US-09-045-973-3
16	28	70.0	269	2	US-07-857-224B-48
17	28	70.0	412	4	US-09-134-001C-4885
18	28	70.0	452	1	US-08-275-488A-2
19	28	70.0	452	1	US-08-275-490-2
20	28	70.0	452	1	US-08-446-380-2
21	28	70.0	452	1	US-08-446-374-2
22	28	70.0	452	1	US-08-446-382-2
23	28	70.0	452	1	US-08-445-801-2
24	28	70.0	452	1	US-08-275-487-2
25	28	70.0	452	5	PCR-US95-08919-2
26	28	70.0	453	1	US-08-275-488A-12
27	28	70.0	453	1	US-08-275-490-12

28	28	70.0	453	1	US-08-446-380-12	Sequence 12, Appl
29	28	70.0	453	1	US-08-446-374-12	Sequence 12, Appl
30	28	70.0	453	1	US-08-446-382-12	Sequence 12, Appl
31	28	70.0	453	1	US-08-445-801-12	Sequence 12, Appl
32	28	70.0	453	1	US-08-275-487-12	Sequence 12, Appl
33	28	70.0	453	5	PCR-US95-08919-12	Sequence 4, Appl
34	28	70.0	494	1	US-08-275-488A-4	Sequence 4, Appl
35	28	70.0	494	1	US-08-275-490-4	Sequence 4, Appl
36	28	70.0	494	1	US-08-446-380-4	Sequence 4, Appl
37	28	70.0	494	1	US-08-446-374-4	Sequence 4, Appl
38	28	70.0	494	1	US-08-446-382-4	Sequence 4, Appl
39	28	70.0	494	1	US-08-445-801-4	Sequence 4, Appl
40	28	70.0	494	1	US-08-275-487-4	Sequence 4, Appl
41	28	70.0	494	5	PCR-US95-08919-4	Sequence 4, Appl
42	28	70.0	609	1	US-08-324-977-40	Sequence 40, Appl
43	28	70.0	609	2	US-08-384-616-40	Sequence 40, Appl
44	28	70.0	609	2	US-08-904-686A-40	Sequence 40, Appl
45	28	70.0	609	4	US-09-315-850-40	Sequence 40, Appl

## ALIGNMENTS

RESULT 1  
US-08-611-977-2  
; Sequence 2, Application US/08611977  
; Patent No. 5972886  
; GENERAL INFORMATION:  
; APPLICANT: TSUJIMOTO, Masafumi  
; APPLICANT: IWASA, Fuyuki  
; APPLICANT: TSUROOKA, No. 5972886uo  
; APPLICANT: NAKAZATO, Hiroshi  
; APPLICANT: MIURA, Kenju  
; APPLICANT: ISHIDA, No. 5972886uhio  
; APPLICANT: KURIHARA, Tatsuya  
; APPLICANT: YAMAGUCHI, Kozo  
; TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
; NUMBER OF SEQUENCES: 34  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: P.O. Box 1404  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/611,977  
; FILING DATE: 06-MAR-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/091,028  
; FILING DATE: 14-JUL-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 4-212305  
; FILING DATE: 17-JUL-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 6-067339  
; FILING DATE: 04-MAR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McGowan, Malcolm K.  
; REGISTRATION NUMBER: 39,300  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:

LENGTH: 7 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-611-977-2

Query Match 100.0%; Score 40; DB 2; Length 7;  
Best Local Similarity 100.0%; Pred. No. 2e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WOSAFTK 7  
1111111  
Db 1 WOSAFTK 7

RESULT 2  
US-08-472-659-34

Sequence 34, Application US/08472659

Patent No. 5831030

GENERAL INFORMATION:

APPLICANT: TSUJIMOTO, Masafumi

APPLICANT: IWASA, Fuyuki

APPLICANT: TSURUOKA, No. 5831030uo

APPLICANT: NAKAZATO, Hiroshi

APPLICANT: MIURA, Kenju

APPLICANT: ISHIDA, No. 5831030unhiro

APPLICANT: KURIHARA, Tatsuya

APPLICANT: YAMAGUCHI, Kozo

APPLICANT: YAMAGUCHI, No. 5831030om1

TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR

NUMBER OF SEQUENCES: 34

CORRESPONDENCE ADDRESS:

ADDRESSEE: Burns, Doane, Swecker & Mathis

STREET: P.O. Box 1404

CITY: Alexandria

STATE: Virginia

COUNTRY: United States

ZIP: 22313-1404

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/472, 659

FILING DATE: 07-JUN-1995

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 4-212305

FILING DATE: 17-JUL-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 5-067339

FILING DATE: 04-MAR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/091,028

FILING DATE: 14-JUL-1993

ATTORNEY/AGENT INFORMATION:

NAME: McGowan, Malcolm K.

REGISTRATION NUMBER: 39,300

REFERENCE/DOCKET NUMBER: 001560-248

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 836-6620

TELEFAX: (703) 836-2021

INFORMATION FOR SEQ ID NO: 34:

SEQUENCE CHARACTERISTICS:

LENGTH: 380 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-472-659-34

Query Match 100.0%; Score 40; DB 2; Length 380;

Best Local Similarity 100.0%; Pred. No. 2e+05;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WOSAFTK 7

1111111

Db 1 WOSAFTK 7

181 WOSAFTK 187

RESULT 4

Query Match 100.0%; Score 40; DB 2; Length 380;

Best Local Similarity 100.0%; Pred. No. 2.3;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WOSAFTK 7

1111111

Db 181 WOSAFTK 187

RESULT 4



US-08-611-977-34  
; Sequence 34, Application US/08611977  
; Patent No. 5972886  
; GENERAL INFORMATION:  
; APPLICANT: TSUTSUMOTO, Masafumi  
; APPLICANT: IWASA, Fuyuki  
; APPLICANT: TSURUOKA, No. 5972886uo  
; APPLICANT: NAKAZATO, Hiroshi  
; APPLICANT: MIURA, Kenju  
; APPLICANT: ISHIDA, No. 5972886uhitro  
; APPLICANT: KURIHARA, Tatsuya  
; APPLICANT: YAMAUCHI, Kozo  
; APPLICANT: YAMAGUCHI, No. 5972886oml  
; TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
; NUMBER OF SEQUENCES: 34  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: P. O. Box 1404  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22133-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/611,977  
; FILING DATE: 06-MAR-1996  
; CLASSIFICATION: 435  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US 08/091,028  
; FILING DATE: 14-JUL-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 4-212305  
; FILING DATE: 17-JUL-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 6-067339  
; FILING DATE: 04-MAR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McGowan, Malcolm K.  
; REGISTRATION NUMBER: 39,300  
; REFERENCE/DOCKET NUMBER: 001560-204  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 34:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 380 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-611-977-34

Query Match 100.0%; Score 40; DB 2; Length 380;  
Best Local Similarity 100.0%; Pred. No. 2.3;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOSAFTK 7  
1111111  
DB 181 MOSAFTK 187

RESULT 5  
US-09-053-866-2  
; Sequence 2, Application US/09053866  
; Patent No. 6111075  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Wenfeng  
; APPLICANT: Presnell, Scott R.  
; APPLICANT: Yee, David P.  
; APPLICANT: Foster, Donald C.

TITLE OF INVENTION: PROTEASE-ACTIVATED RECEPTOR  
; TITLE OF INVENTION: PAR4 (ZCHEMR2)  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Zymogenetics, Inc.  
; STREET: 1201 Eastlake Avenue East  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98102  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/053,866  
; FILING DATE:  
; CLASSIFICATION:  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Leith, Debra K  
; REGISTRATION NUMBER: 32,619  
; REFERENCE/DOCKET NUMBER: 98-10  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 206-442-6674  
; TELEFAX: 206-442-6678  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 385 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
; US-09-053-866-2

Query Match 75.0%; Score 30; DB 3; Length 385;  
Best Local Similarity 83.3%; Pred. No. 1.8e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MOSAFT 6  
111111  
DB 241 MOSAFT 246

RESULT 6  
US-09-479-130-2  
; Sequence 2, Application US/09479130  
; Patent No. 6436400  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Wenfeng  
; APPLICANT: Presnell, Scott R.  
; APPLICANT: Yee, David P.  
; APPLICANT: Foster, Donald C.  
; TITLE OF INVENTION: PROTEASE-ACTIVATED RECEPTOR  
; TITLE OF INVENTION: PAR4 (ZCHEMR2)  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Zymogenetics, Inc.  
; STREET: 1201 Eastlake Avenue East  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98102  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/479,130  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Leitch, Debra K  
REGISTRATION NUMBER: 32,619  
REFERENCE/DOCKET NUMBER: 98-10  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-442-6674  
TELEFAX: 206-442-6678  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 385 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
US-09-479-130-2

Query Match 75.0%; Score 30; DB 4; Length 385;  
Best Local Similarity 83.3%; Pred. No. 1.8e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 WOSAFT 6  
111111

Db 241 WOSAFT 246

RESULT 7  
US-09-216-295-22  
Sequence 22, Application US/09216295  
Patent No. 6268328  
GENERAL INFORMATION:  
APPLICANT: Mitchinson, Colin  
APPLICANT: Mendt, Dan J.  
TITLE OF INVENTION: No. 6268328el Variant Egitr-Like Cellulase Compositions  
FILE REFERENCE: GC555  
CURRENT APPLICATION NUMBER: US/09/216,295  
CURRENT FILING DATE: 1998-12-18  
NUMBER OF SEQ ID NOS: 41  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 22  
LENGTH: 381  
TYPE: PRT  
ORGANISM: Streptomyces lividans CelB  
US-09-216-295-22

Query Match 72.5%; Score 29; DB 4; Length 381;  
Best Local Similarity 83.3%; Pred. No. 2.8e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 WOSAFT 6  
111111

Db 309 WOLAFT 314

RESULT 8  
US-08-543-238-11  
Sequence 11, Application US/08543238  
Patent No. 5607919  
GENERAL INFORMATION:  
APPLICANT: Bojsen, Kirsten  
APPLICANT: Kragh, Karsten M.  
APPLICANT: Mikkelsen, Jorn D.  
APPLICANT: Nielsen, Klaus K.  
TITLE OF INVENTION: Anti-Microbial Proteins  
NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sandoz Agro, Inc.  
STREET: 975 California Avenue  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/543,238  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Marcus-Wyner, Lynn  
REGISTRATION NUMBER: 34,869  
REFERENCE/DOCKET NUMBER: 137-1078/MA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/354-3588  
TELEFAX: 415/857-1125  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 146 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-543-238-11

Query Match 70.0%; Score 28; DB 1; Length 146;  
Best Local Similarity 57.1%; Pred. No. 1.7e+02;  
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 WOSAFT 7  
111111

Db 116 WDFVFTK 122

RESULT 9  
US-08-420-526-11  
Sequence 11, Application US/08420526  
Patent No. 5608151  
GENERAL INFORMATION:  
APPLICANT: Bojsen, Kirsten  
APPLICANT: Kragh, Karsten M.  
APPLICANT: Mikkelsen, Jorn D.  
APPLICANT: Nielsen, Klaus K.  
TITLE OF INVENTION: Anti-Microbial Proteins  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sandoz Agro, Inc.  
STREET: 975 California Avenue  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/420,526  
FILING DATE:  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Marcus-Wyner, Lynn  
REGISTRATION NUMBER: 34,869  
REFERENCE/DOCKET NUMBER: 137-1078/MA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/354-3588

TELEFAX: 415/857-1125  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 146 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-420-526-11

Query Match  
Best Local Similarity 70.0%; Score 28; DB 1; Length 146;  
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 WSAFT 7  
1 1 1 1 1  
DB 116 MDVFTK 122

RESULT 10  
US-09-134-001C-4482  
Sequence 4482, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 4482  
LENGTH: 177  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4482

Query Match  
Best Local Similarity 70.0%; Score 28; DB 4; Length 177;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 WSAFT 6  
1 1 1 1 1  
DB 88 WSAFT 93

RESULT 11  
US-08-483-695-46  
Sequence 46, Application US/08483695  
Patent No. 5866139  
GENERAL INFORMATION:  
APPLICANT: Brechot, Christian  
APPLICANT: Kremsdorf, Dina  
APPLICANT: Porchon, Colette  
TITLE OF INVENTION: Nucleotide and Peptide Sequences of a  
TITLE OF INVENTION: Hepatitis C Virus Isolate, Diagnostic and Therapeutic  
NUMBER OF SEQUENCES: 46  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
ADDRESSEE: Dunner  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/483,695  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/965,285  
FILING DATE: 18-MAR-1993  
APPLICATION NUMBER: FR 91 06 882  
FILING DATE: 06-JUN-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J.  
REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 05286-0001-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400

Query Match  
Best Local Similarity 70.0%; Score 28; DB 2; Length 189;  
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WSAFT 6  
1 1 1 1 1  
DB 104 WSAFT 109

RESULT 12  
US-07-965-285-46  
Sequence 46, Application US/07965285  
Patent No. 5879904  
GENERAL INFORMATION:  
APPLICANT: Brechot, Christian  
APPLICANT: Kremsdorf, Dina  
APPLICANT: Porchon, Colette  
TITLE OF INVENTION: Nucleotide and Peptide Sequences of a  
TITLE OF INVENTION: Hepatitis C Virus Isolate, Diagnostic and Therapeutic  
NUMBER OF SEQUENCES: 46  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
ADDRESSEE: Dunner  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/965,285  
FILING DATE: 18-MAR-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 91 06 882  
FILING DATE: 06-JUN-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J.  
REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 05286-0001-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400

INFORMATION FOR SEQ ID NO: 46:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 189 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-07-965-285-46

Query Match 70.0%; Score 28; DB 2; Length 189;  
Best Local Similarity 66.7%; Pred. No. 2.2e+02;  
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 WSAFT 6  
DB 104 WESVFT 109

RESULT 13  
US-08-487-231-46  
Sequence 46, Application US/08487231  
Patent No. 5919454  
GENERAL INFORMATION:  
APPLICANT: Brechot, Christian  
APPLICANT: Kremsdorf, Dina  
TITLE OF INVENTION: Nucleotide and Peptide Sequences of a  
TITLE OF INVENTION: Hepatitis C Virus Isolate, Diagnostic and Therapeutic  
NUMBER OF SEQUENCES: 46  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
ADDRESS: Dunner  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,231  
FILING DATE: 07-JUNE-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/965,285  
FILING DATE: 18-MAR-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 91 06 882  
FILING DATE: 06-JUN-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J.  
REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 05286-0001-02000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 46:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 189 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-487-231-46

Query Match 70.0%; Score 28; DB 2; Length 189;  
Best Local Similarity 66.7%; Pred. No. 2.2e+02;  
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
OY 1 WSAFT 6

DB 104 WESVFT 109

RESULT 14  
US-09-201-912-46  
Sequence 46, Application US/09201912  
Patent No. 6210962  
GENERAL INFORMATION:  
APPLICANT: Brechot, Christian  
APPLICANT: Kremsdorf, Dina  
TITLE OF INVENTION: Nucleotide and Peptide Sequences of a  
TITLE OF INVENTION: Hepatitis C Virus Isolate, Diagnostic and Therapeutic  
NUMBER OF SEQUENCES: 46  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
ADDRESS: Dunner  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/201,912  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/965,285  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J.  
REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 05286-0001-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 46:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 189 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-201-912-46

Query Match 70.0%; Score 28; DB 4; Length 189;  
Best Local Similarity 66.7%; Pred. No. 2.2e+02;  
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 WSAFT 6  
DB 104 WESVFT 109

RESULT 15  
US-09-045-973-3  
Sequence 3, Application US/09045973  
Patent No. 6165767  
GENERAL INFORMATION:  
APPLICANT: Lal, Preeti  
APPLICANT: Yue, Henry  
APPLICANT: Corley, Neil C.  
APPLICANT: Guegler, Karl J.  
APPLICANT: Baughn, Mariah  
TITLE OF INVENTION: PROTEIN PHOSPHATASE RELATED MOLECULES  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/045,973  
FILING DATE: Filed Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0491 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 261 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: BRAINOT18  
CLONE: 2534680  
US-09-045-973-3

Query Match 70.0%; Score 28; DB 4; Length 261;

Best Local Similarity 66.7%; Pred. No. 3e+02; Mismatches 0; Indels 0; Gaps 0;

QY 1 WOSAF 6  
DB 118 WRAFT 123

Search completed: July 11, 2003, 12:02:05  
Job time : 2.25156 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:59:17 ; Search time 2.05198 Seconds

(without alignments)  
397.148 Million cell updates/sec

Title: US-10-091-442-2

Perfect score: 40

Sequence: 1 MOSAFWK 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 445758 seqs, 116419773 residues

Total number of hits satisfying chosen parameters: 445758

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications\_AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_NEM\_PUB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PC1\_NEM\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEM\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/PC1US\_PUBCOMB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEM\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEM\_PUB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEM\_PUB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEM\_PUB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	100.0	7	9	US-10-091-442-2
2	40	100.0	7	9	US-10-140-719-2
3	40	100.0	380	9	US-10-091-442-34
4	40	100.0	380	10	US-09-140-719-34
5	31	77.5	125	9	US-09-738-626-5215
6	31	77.5	168	10	US-09-764-853-656
7	31	77.5	194	9	US-10-091-438-157
8	31	77.5	198	12	US-10-007-693-65
9	30	75.0	195	9	US-10-097-065-172
10	30	75.0	385	9	US-10-225-567A-516
11	30	75.0	687	9	US-09-738-626-6127
12	29	72.5	145	9	US-10-204-887-158
13	29	72.5	165	9	US-09-738-626-6151
14	29	72.5	225	9	US-09-738-626-4605
15	29	72.5	307	9	US-09-854-133-397
16	29	72.5	307	9	US-10-144-649A-397
17	29	72.5	307	10	US-09-738-973-397
18	29	72.5	576	9	US-09-364-847-37
19	29	72.5	693	9	US-10-290-078-9

20	29	72.5	712	9	US-09-364-847-49	Sequence 49, Appl
21	29	72.5	712	9	US-09-364-847-51	Sequence 51, Appl
22	29	72.5	1203	9	US-10-097-340-43	Sequence 43, Appl
23	28	70.0	95	9	US-09-899-046-30	Sequence 30, Appl
24	28	70.0	95	9	US-09-878-281-30	Sequence 30, Appl
25	28	70.0	131	10	US-09-814-122-70	Sequence 70, Appl
26	28	70.0	209	9	US-09-899-046-223	Sequence 223, App
27	28	70.0	209	9	US-09-878-281-223	Sequence 223, App
28	28	70.0	224	9	US-10-156-761-7796	Sequence 7796, Ap
29	28	70.0	243	9	US-09-738-626-6003	Sequence 6003, Ap
30	28	70.0	382	9	US-10-078-770-66	Sequence 66, Appl
31	28	70.0	429	10	US-09-815-242-14022	Sequence 14022, A
32	28	70.0	461	9	US-09-925-299-957	Sequence 957, App
33	28	70.0	461	9	US-10-106-658-5254	Sequence 5254, Ap
34	28	70.0	461	10	US-09-925-299-957	Sequence 957, App
35	28	70.0	503	9	US-10-142-231-63	Sequence 63, Appl
36	28	70.0	631	9	US-10-214-932-78	Sequence 78, Appl
37	28	70.0	768	9	US-10-156-761-13830	Sequence 13830, A
38	28	70.0	1692	9	US-09-919-901-4	Sequence 4, Appl1
39	28	70.0	1692	9	US-09-919-901-11	Sequence 11, Appl
40	28	70.0	1692	9	US-09-919-901-18	Sequence 18, Appl
41	28	70.0	1985	9	US-10-259-275-42	Sequence 42, Appl
42	28	70.0	2201	9	US-10-085-476-2	Sequence 2, Appl1
43	28	70.0	2201	12	US-10-029-907-3	Sequence 3, Appl1
44	28	70.0	2307	9	US-09-919-901-2	Sequence 2, Appl1
45	28	70.0	2307	9	US-09-919-901-9	Sequence 9, Appl1

#### ALIGNMENTS

RESULT 1  
US-10-091-442-2  
; Sequence 2, Application US/10091442  
; Patent No. US20020164711A1  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
IMASA, Fuyuki  
TSURUOKA, No. US20020164711A1uo  
NAKAZATO, Hiroshi  
MIURA, Kenju  
ISHIDA, No. US20020164711A1uhfro  
KURIHARA, Tatsuya  
YAMAGUCHI, Kozo  
YAMAGUCHI, No. US20020164711A1omf  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/091,442  
FILING DATE: 07-Mar-2002  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-213305  
FILING DATE: 17-JUL-1992  
APPLICATION NUMBER: JP 6-067339

FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-10-091-442-2

Query Match 100.0%; Score 40; DB 9; Length 7;  
Best Local Similarity 100.0%; Pred. No. 4e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WOSAFK 7  
111111  
DB 1 WOSAFK 7

RESULT 2  
US-09-140-719-2  
Sequence 2, Application US/09140719  
Patent No. US20010026931A1  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSUROUOKA, No. US20010026931A1uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. US20010026931A1uhitro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAGUCHI, No. US20010026931A1omi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION NUMBER:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.

REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-140-719-2

Query Match 100.0%; Score 40; DB 10; Length 7;  
Best Local Similarity 100.0%; Pred. No. 4e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WOSAFK 7  
111111  
DB 1 WOSAFK 7

RESULT 3  
US-10-091-442-34  
Sequence 34, Application US/10091442  
Patent No. US20020164711A1  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSUROUOKA, No. US20020164711A1uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. US20020164711A1uhitro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAGUCHI, No. US20020164711A1omi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/091,442  
FILING DATE: 07-Mar-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021



INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 34:  
US-10-091-442-34

Query Match 100.0%; Score 40; DB 9; Length 380;  
Best Local Similarity 100.0%; Pred. No. 3.6;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WSAFTK 7  
1111111  
DB 181 WSAFTK 187

## RESULT 4

US-09-140-719-34  
Sequence 34, Application US/09140719  
Patent No. US20010026931A1  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSURUOKA, No. US20010026931A1no  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. US20010026931A1nhiro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAUCHI, Kozo  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid

TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-140-719-34

Query Match 100.0%; Score 40; DB 10; Length 380;  
Best Local Similarity 100.0%; Pred. No. 3.6;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WSAFTK 7  
1111111  
DB 181 WSAFTK 187

## RESULT 5

US-09-738-626-5215  
Sequence 5215, Application US/09738626  
Publication No. US20020197605A1  
GENERAL INFORMATION:  
APPLICANT: NAKAGAWA, SATOSHI  
APPLICANT: MIZOGUCHI, HIROSHI  
APPLICANT: ANDO, SEIKO  
APPLICANT: HAYASHI, MIKIRO  
APPLICANT: OCHIAI, KEIKO  
APPLICANT: YOKOI, HARUHIKO  
APPLICANT: TATEISHI, NAOKO  
APPLICANT: SENOH, AKIHIRO  
APPLICANT: IKEDA, MASATO  
APPLICANT: OKAZAKI, AKIO  
TITLE OF INVENTION: NOVEL POLYPEPTIDES  
FILE REFERENCE: 249-125  
CURRENT APPLICATION NUMBER: US/09/738,626  
CURRENT FILING DATE: 2000-12-18  
PRIOR APPLICATION NUMBER: JP 99/377484  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: JP 00/159162  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: JP 00/280988  
PRIOR FILING DATE: 2000-08-03  
NUMBER OF SEQ ID NOS: 7059  
SOFTWARE: PatentIn ver. 3.0  
SEQ ID NO 5215  
LENGTH: 125  
TYPE: PRT  
ORGANISM: Corynebacterium glutamicum  
US-09-738-626-5215

Query Match 77.5%; Score 31; DB 9; Length 125;  
Best Local Similarity 71.4%; Pred. No. 67;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WSAFTK 7  
11:111  
DB 30 WQNALTK 36

RESULT 6  
US-09-764-853-656  
Sequence 656, Application US/09764853  
Patent No. US20020090672A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: P3206  
CURRENT APPLICATION NUMBER: US/09/764,853  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 939  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 656  
LENGTH: 168  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-764-853-656

Query Match 77.5% Score 31; DB 10; Length 168;  
Best Local Similarity 71.4%; Pred. No. 90;  
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 WSAFTK 7  
DB 49 WQSVENK 55

## RESULT 7

US-10-091-438-157  
Sequence 157, Application US/10091438  
Publication No. US20030077606A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PT217C1  
CURRENT APPLICATION NUMBER: US/10/091,438  
CURRENT FILING DATE: 2001-01-17  
PRIOR APPLICATION NUMBER: 09/764,879  
PRIOR FILING DATE: 2001-01-17  
PRIOR APPLICATION NUMBER: 60/179,065  
PRIOR FILING DATE: 2000-01-31  
PRIOR APPLICATION NUMBER: 60/180,628  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: 60/214,886  
PRIOR FILING DATE: 2000-06-28  
PRIOR APPLICATION NUMBER: 60/217,487  
PRIOR FILING DATE: 2000-07-11  
PRIOR APPLICATION NUMBER: 60/225,758  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/220,963  
PRIOR FILING DATE: 2000-07-26  
PRIOR APPLICATION NUMBER: 60/217,496  
PRIOR FILING DATE: 2000-07-11  
PRIOR APPLICATION NUMBER: 60/225,447  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/218,290  
PRIOR FILING DATE: 2000-07-14  
PRIOR APPLICATION NUMBER: 60/225,757  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/226,868  
PRIOR FILING DATE: 2000-08-22  
PRIOR APPLICATION NUMBER: 60/216,647  
PRIOR FILING DATE: 2000-07-07  
PRIOR APPLICATION NUMBER: 60/225,267  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/216,880  
PRIOR FILING DATE: 2000-07-07  
PRIOR APPLICATION NUMBER: 60/225,270  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/251,869  
PRIOR FILING DATE: 2000-12-08  
PRIOR APPLICATION NUMBER: 60/235,834  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: 60/234,274  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: 60/234,223  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: 60/228,924  
PRIOR FILING DATE: 2000-08-30  
PRIOR APPLICATION NUMBER: 60/224,518  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/236,369  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/224,519  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/220,964  
PRIOR FILING DATE: 2000-07-26  
PRIOR APPLICATION NUMBER: 60/241,809  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/249,299

PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/236,327  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/241,785  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/244,617  
PRIOR FILING DATE: 2000-11-01  
PRIOR APPLICATION NUMBER: 60/225,268  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/236,368  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/251,856  
PRIOR FILING DATE: 2000-12-08  
PRIOR APPLICATION NUMBER: 60/251,868  
PRIOR FILING DATE: 2000-12-08  
PRIOR APPLICATION NUMBER: 60/229,344  
PRIOR FILING DATE: 2000-09-01  
PRIOR APPLICATION NUMBER: 60/234,997  
PRIOR FILING DATE: 2000-09-25  
PRIOR APPLICATION NUMBER: 60/229,343  
PRIOR FILING DATE: 2000-09-01  
PRIOR APPLICATION NUMBER: 60/229,345  
PRIOR FILING DATE: 2000-09-01  
PRIOR APPLICATION NUMBER: 60/229,287  
PRIOR FILING DATE: 2000-09-01  
PRIOR APPLICATION NUMBER: 60/229,513  
PRIOR FILING DATE: 2000-09-05  
PRIOR APPLICATION NUMBER: 60/231,413  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/229,509  
PRIOR FILING DATE: 2000-09-05  
PRIOR APPLICATION NUMBER: 60/236,367  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/237,039  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/237,038  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/236,370  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/236,802  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/237,037  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/237,040  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/240,960  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/239,935  
PRIOR FILING DATE: 2000-10-13  
PRIOR APPLICATION NUMBER: 60/239,937  
PRIOR FILING DATE: 2000-10-13  
PRIOR APPLICATION NUMBER: 60/241,787  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/246,474  
PRIOR FILING DATE: 2000-11-08  
PRIOR APPLICATION NUMBER: 60/246,532  
PRIOR FILING DATE: 2000-11-08  
PRIOR APPLICATION NUMBER: 60/249,216  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,210  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/226,681  
PRIOR FILING DATE: 2000-08-22  
PRIOR APPLICATION NUMBER: 60/225,759  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/225,213  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/227,182  
PRIOR FILING DATE: 2000-08-22  
PRIOR APPLICATION NUMBER: 60/225,214  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/235,836  
PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: 60/230,438  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/215,135  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: 60/225,266  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/249,218  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,208  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,213  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,212  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,207  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,245  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,244  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,217  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,211  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,215  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,264  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,214  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,297  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/232,400  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/231,242  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/232,081  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/232,080  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/231,414  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/231,244  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/233,064  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/233,063  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/232,397  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/232,399  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/232,401  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/241,808  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/241,826  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/241,786  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/241,221  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/246,475  
PRIOR FILING DATE: 2000-11-06  
PRIOR APPLICATION NUMBER: 60/231,243  
PRIOR FILING DATE: 2000-09-06

Query Match 77.5% Score 31: DB 9: Length 194:

Best Local Similarity 71.4% Pred. No. 1e+02: 2: Indels 0: Gaps 0:

Matches 5: Conservative 0: Mismatches

QY 1 WSAFTK 7  
||| |

Db 166 WQSVFNK 172  
RESULT 8  
US-10-007-693-65  
Sequence 65, Application US/10007693  
Patent No. US20020146776A1  
GENERAL INFORMATION:  
APPLICANT: Bhatia, Ajay  
APPLICANT: Probst, Peter  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT  
FILE REFERENCE: 210121.515C2  
CURRENT APPLICATION NUMBER: US/10/007,693  
NUMBER OF SEQ ID NOS: 157  
SEQ ID NO 65  
LENGTH: 978  
TYPE: PRT  
ORGANISM: Chlamydia pneumoniae  
US-10-007-693-65

Query Match 77.5% Score 31: DB 12: Length 978:  
Best Local Similarity 71.4% Pred. No. 5.2e+02:  
Matches 5: Conservative 1: Mismatches 1: Indels 0: Gaps 0:

QY 1 WSAFTK 7  
||| |  
Db 854 WQSKFTE 860

RESULT 9  
US-10-097-065-172  
Sequence 172, Application US/10097065  
Publication No. US20030055236A1  
GENERAL INFORMATION:  
APPLICANT: Moore, Paul A. et al.  
TITLE OF INVENTION: 110 Human Secreted Proteins  
FILE REFERENCE: P2021P1  
CURRENT APPLICATION NUMBER: US/10/097,065  
CURRENT FILING DATE: 2002-03-14  
PRIOR APPLICATION NUMBER: PCT/US98/27059  
PRIOR FILING DATE: 1998-12-17  
PRIOR APPLICATION NUMBER: 60/070,923  
PRIOR FILING DATE: 1997-12-18  
PRIOR APPLICATION NUMBER: 60/068,007  
PRIOR FILING DATE: 1997-12-18  
PRIOR APPLICATION NUMBER: 60/068,057  
PRIOR FILING DATE: 1997-12-18  
PRIOR APPLICATION NUMBER: 60/068,006  
PRIOR FILING DATE: 1997-12-18  
PRIOR APPLICATION NUMBER: 60/068,369  
PRIOR FILING DATE: 1997-12-19  
PRIOR APPLICATION NUMBER: 60/068,367  
PRIOR FILING DATE: 1997-12-19  
PRIOR APPLICATION NUMBER: 60/068,368  
PRIOR FILING DATE: 1997-12-19  
PRIOR APPLICATION NUMBER: 60/068,169  
PRIOR FILING DATE: 1997-12-19  
PRIOR APPLICATION NUMBER: 60/068,053  
PRIOR FILING DATE: 1997-12-18  
PRIOR APPLICATION NUMBER: 60/068,064  
PRIOR FILING DATE: 1997-12-18  
PRIOR APPLICATION NUMBER: 60/068,054  
PRIOR FILING DATE: 1997-12-18  
PRIOR APPLICATION NUMBER: 60/068,008  
PRIOR FILING DATE: 1997-12-18  
PRIOR APPLICATION NUMBER: 60/068,365  
PRIOR FILING DATE: 1997-12-19  
NUMBER OF SEQ ID NOS: 672  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 172  
LENGTH: 195

;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: SITE  
;; LOCATION: (195)  
;; OTHER INFORMATION: Xaa equals stop translation  
US-10-097-065-172

Query Match 75.0%; Score 30; DB 9; Length 195;  
Best Local Similarity 71.4%; Pred. No. 1.6e+02;  
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 WQSAFTK 7  
| | | | |  
Db 178 WMSMFTK 184

RESULT 10  
US-10-225-567A-516  
; Sequence 516, Application US/10225567A  
; Publication No. US20030113798A1  
; GENERAL INFORMATION:  
; APPLICANT: Lifespan Biosciences  
; APPLICANT: Brown, Joseph P.  
; APPLICANT: Burner, Glenna C.  
; APPLICANT: Roush, Christine L.  
; TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTORS  
; FILE REFERENCE: 1920-4-4  
; CURRENT APPLICATION NUMBER: US/10/225,567A  
; CURRENT FILING DATE: 2001-12-19  
; PRIOR APPLICATION NUMBER: 60/257,144  
; PRIOR FILING DATE: 2000-12-19  
; NUMBER OF SEQ ID NOS: 2292  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 516  
; LENGTH: 385  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-225-567A-516

Query Match 75.0%; Score 30; DB 9; Length 385;  
Best Local Similarity 83.3%; Pred. No. 3.2e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 WQSAFT 6  
| | | | |  
Db 241 WQPAFT 246

RESULT 11  
US-09-738-626-6127  
; Sequence 6127, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: MAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988

;; PRIOR FILING DATE: 2000-08-03  
;; NUMBER OF SEQ ID NOS: 7059  
;; SOFTWARE: PatentIn ver. 3.0  
;; SEQ ID NO 6127  
;; LENGTH: 687  
;; TYPE: PRT  
;; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-6127

Query Match 75.0%; Score 30; DB 9; Length 687;  
Best Local Similarity 100.0%; Pred. No. 5.8e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WQSAF 5  
| | | | |  
Db 377 WQSAF 381

RESULT 12  
US-10-204-887-158  
; Sequence 158, Application US/10204887  
; Publication No. US20030124569A1  
; GENERAL INFORMATION:  
; APPLICANT: INCYTE GENOMICS, INC.  
; APPLICANT: PANZER, Scott R.  
; APPLICANT: SPIRO, Peter A.  
; APPLICANT: BANVILLE, Steven C.  
; APPLICANT: SHAH, Purvi  
; APPLICANT: CHALUP, Michael S.  
; APPLICANT: CHANG, Simon C.  
; APPLICANT: CHEN, Alice  
; APPLICANT: D'SA, Steven A.  
; APPLICANT: AMSHEY, Stefan  
; APPLICANT: DAHL, Christopher R.  
; APPLICANT: DAM, Tam C.  
; APPLICANT: DANIELS, Susan E.  
; APPLICANT: DOFOUR, Gerard E.  
; APPLICANT: FLORES, Vincent  
; APPLICANT: FONG, Willy T.  
; APPLICANT: GREENMALT, Lila B.  
; APPLICANT: HITLMAN, Jennifer L.  
; APPLICANT: JONES, Anissa L.  
; APPLICANT: LIU, Tommy F.  
; APPLICANT: ROSEBERRY, Ann M.  
; APPLICANT: ROSEN, Bruce H.  
; APPLICANT: RUSSO, Frank D.  
; APPLICANT: STOCKREHER, Theresa K.  
; APPLICANT: DAFEO, Abel  
; APPLICANT: WRIGHT, Rachel J.  
; APPLICANT: VAP, Pierre E.  
; APPLICANT: YU, Jimmy Y.  
; APPLICANT: BRADLEY, Diana L.  
; APPLICANT: BRATCHER, Shawn R.  
; APPLICANT: CHEN, Wensheng  
; APPLICANT: COHEN, Howard J.  
; APPLICANT: HODGSON, David M.  
; APPLICANT: LINCOLN, Stephen E.  
; TITLE OF INVENTION: SECRETORY MOLECULES  
; FILE REFERENCE: PT-1134 PCT  
; CURRENT APPLICATION NUMBER: US/10/204,887  
; CURRENT FILING DATE: 2002-08-21  
; PRIOR APPLICATION NUMBER: 60/185,215; 60/185,216; 60/205,232; 60/205,323; 60/205,2  
60/205,324; 60/205,286  
; PRIOR FILING DATE: 2000-02-24; 2000-02-24; 2000-05-16; 2000-05-17; 2000-05-17;  
2000-05-17; 2000-05-17  
; NUMBER OF SEQ ID NOS: 159  
; SOFTWARE: PERL Program  
; SEQ ID NO 158  
; LENGTH: 145  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature

OTHER INFORMATION: Incyte ID No. US20030124569A1 LI:221836.3.orf2:2000WAY01  
US-10-204-887-158

Query Match 72.5%; Score 29; DB 9; Length 145;  
Best Local Similarity 71.4%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 WQSAFTK 7  
111111  
DB 111 WQSEFK 117

RESULT 13  
US-09-738-626-6151

Sequence 6151, Application US/09738626  
Publication No. US20020197605A1

GENERAL INFORMATION:

APPLICANT: NAKAGAWA, SATOSHI  
APPLICANT: MIZOGUCHI, HIROSHI

APPLICANT: ANDO, SEIKO  
APPLICANT: HAYASHI, MIKIRO

APPLICANT: OCHIAI, KEIKO  
APPLICANT: YOKOI, HARUHIKO

APPLICANT: TATEISHI, NAOKO  
APPLICANT: SENOH, AKIHIRO

APPLICANT: IKEDA, MASATO  
APPLICANT: OZAKI, AKIO

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-125

CURRENT APPLICATION NUMBER: US/09/738, 626  
CURRENT FILING DATE: 2000-12-18

PRIOR APPLICATION NUMBER: JP 99/377484  
PRIOR FILING DATE: 1999-12-16

PRIOR APPLICATION NUMBER: JP 00/159162  
PRIOR FILING DATE: 2000-04-07

PRIOR APPLICATION NUMBER: JP 00/280988  
PRIOR FILING DATE: 2000-08-03

NUMBER OF SEQ ID NOS: 7059  
SOFTWARE: PatentIn ver. 3.0

SEQ ID NO 6151  
LENGTH: 165

TYPE: PRT  
ORGANISM: Corynebacterium glutamicum

US-09-738-626-6151

Query Match 72.5%; Score 29; DB 9; Length 165;  
Best Local Similarity 57.1%; Pred. No. 2.2e+02;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 WQSAFTK 7  
111111  
DB 80 WDSFTR 86

RESULT 14  
US-09-738-626-4605

Sequence 4605, Application US/09738626  
Publication No. US20020197605A1

GENERAL INFORMATION:

APPLICANT: NAKAGAWA, SATOSHI  
APPLICANT: MIZOGUCHI, HIROSHI

APPLICANT: ANDO, SEIKO  
APPLICANT: HAYASHI, MIKIRO

APPLICANT: OCHIAI, KEIKO  
APPLICANT: YOKOI, HARUHIKO

APPLICANT: TATEISHI, NAOKO  
APPLICANT: SENOH, AKIHIRO

APPLICANT: IKEDA, MASATO  
APPLICANT: OZAKI, AKIO

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-125

CURRENT APPLICATION NUMBER: US/09/738, 626  
CURRENT FILING DATE: 2000-12-18

PRIOR APPLICATION NUMBER: JP 99/377484  
PRIOR FILING DATE: 1999-12-16

PRIOR APPLICATION NUMBER: JP 00/159162  
PRIOR FILING DATE: 2000-04-07

PRIOR APPLICATION NUMBER: JP 00/280988  
PRIOR FILING DATE: 2000-08-03

NUMBER OF SEQ ID NOS: 7059  
SOFTWARE: PatentIn ver. 3.0

SEQ ID NO 4605  
LENGTH: 225

TYPE: PRT  
ORGANISM: Corynebacterium glutamicum

US-09-738-626-4605

Query Match 72.5%; Score 29; DB 9; Length 225;  
Best Local Similarity 57.1%; Pred. No. 3e+02;  
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 WQSAFTK 7  
111111  
DB 187 WQSAFSE 193

RESULT 15  
US-09-854-133-397

Sequence 397, Application US/09854133  
Publication No. US20020183499A1

GENERAL INFORMATION:

APPLICANT: Lodes, Michael J.  
APPLICANT: Mohamath, Raedoh

APPLICANT: Henderson, Robert A.  
APPLICANT: Benson, Darin R.

APPLICANT: Secrist, Heather  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR

TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
FILE REFERENCE: 210121.475C10

CURRENT APPLICATION NUMBER: US/09/854, 133  
CURRENT FILING DATE: 2001-05-11

NUMBER OF SEQ ID NOS: 735  
SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 397  
LENGTH: 307

TYPE: PRT  
ORGANISM: Homo sapiens

US-09-854-133-397

Query Match 72.5%; Score 29; DB 9; Length 307;  
Best Local Similarity 83.3%; Pred. No. 4e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 WQSAFT 6  
111111  
DB 212 WQVAF 217

Search completed: July 11, 2003, 12:37:32  
Job time : 3.05198 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:45:06 ; Search time 1.41164 Seconds  
(without alignments)  
476.708 Million cell updates/sec

Title: US-10-091-442-2

Perfect score: 40  
Sequence: 1 WQSAFTK 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :  
1: PIR\_73:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	82.5	397	2 F83721	chromate transport
2	33	82.5	865	2 T40288	hypothetical prote
3	33	82.5	880	2 B45956	beta-glucosidase (
4	32	80.0	77	2 S04795	mobilization prote
5	32	80.0	736	2 C69451	cationic amino aci
6	32	80.0	837	2 AD0159	outer membrane ush
7	32	80.0	938	2 F86548	polymorphic outer
8	32	80.0	938	2 H72074	polymorphic membra
9	31	77.5	112	2 H82752	hypothetical prote
10	31	77.5	289	2 H97480	hypothetical prote
11	31	77.5	289	2 AH2698	permease (imported
12	31	77.5	291	2 T64069	hypothetical prote
13	31	77.5	357	2 G81185	hypothetical prote
14	31	77.5	357	2 F81916	probable integral
15	31	77.5	427	2 T32652	hypothetical prote
16	31	77.5	514	2 E75142	endoglucanase PA80
17	31	77.5	590	2 F95853	probable phospholi
18	31	77.5	677	2 S57599	FUS2 protein - yea
19	31	77.5	722	2 B86583	transcription elon
20	31	77.5	722	2 C72040	transcription elon
21	31	77.5	978	2 G72076	polymorphic membra
22	31	77.5	978	2 B81593	polymorphic membra
23	31	77.5	978	2 C86547	polymorphic outer
24	31	77.5	1108	2 AP1271	DNA polymerase III
25	31	77.5	1145	2 A59251	myosin - Acetabula
26	30	75.0	128	2 H97098	probable cation ef
27	30	75.0	165	2 D69784	hypothetical prote
28	30	75.0	261	2 T16550	hypothetical prote
29	30	75.0	266	2 T16106	hypothetical prote

30	30	75.0	273	2 AF2025	hypothetical prote
31	30	75.0	295	2 S75994	hypothetical prote
32	30	75.0	266	2 S72601	hypothetical prote
33	30	75.0	317	2 D84098	hypothetical prote
34	30	75.0	324	2 C95864	conserved hypotnet
35	30	75.0	337	2 G97742	cytochrome c oxida
36	30	75.0	351	1 XUBPB4	DNA beta-glucosylt
37	30	75.0	359	2 D88940	protein C05E4.1 (1
38	30	75.0	415	2 AG0301	probable membrane
39	30	75.0	429	1 AUPBOD	glutamate-amonnia
40	30	75.0	432	2 T14292	glutamate-amonnia
41	30	75.0	479	2 T20700	hypothetical prote
42	30	75.0	494	2 E70352	anthranilate synth
43	30	75.0	504	2 S75597	hypothetical prote
44	30	75.0	521	2 T49355	related to protein
45	30	75.0	550	2 T40379	hypothetical prote

## ALIGNMENTS

RESULT 1  
F83721  
chromate transporter chra [imported] - Bacillus halodurans (strain C-125)  
C:Species: Bacillus halodurans  
C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
C:Accession: F83721  
R:Takami, H.; Nakasone, K.; Takaki, Y.; Mieno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H  
Nucleic Acids Res. 28, 4317-4331, 2000  
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a  
A:Reference number: A83650; MUID:20512582; PMID:11058132  
A:Accession: F83721  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-397 <STO>  
A:Cross-references: GB:AP001509; GB:BA000004; NID:g10173176; PIDN:BAR04293.1; GSPDB:G  
A:Experimental source: strain C-125  
C:Genetics:  
A:Gene: chra  
C:Superfamily: chromate resistance protein A

Query Match  
Best Local Similarity 82.5%; Score 33; DB 2; Length 397;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WQSAFTK 7  
II:III:  
Db 160 WQTAFQ 166

RESULT 2  
T40288  
hypothetical protein SPBC354.08c - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: T40288  
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Hubbard, H.; Duesterhoeft, A.  
submitted to the EMBL Data Library, March 1998  
A:Reference number: Z21918  
A:Accession: T40288  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-865 <MOO>  
A:Cross-references: EMBL:AL022071; PIDN:CAA17808.1; GSPDB:GN000067; SPDB:SPBC354.08c  
A:Experimental source: strain 972h-; cosmid c354  
C:Genetics:  
A:Gene: SPDB:SPBC354.08c  
A:Map position: 2  
A:Introns: 58/2

Query Match  
Best Local Similarity 82.5%; Score 33; DB 2; Length 865;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 WOSAFK 7  
11:1111  
DB 448 WOSAFK 454

## RESULT 3

B45956  
beta-glucosidase (EC 3.2.1.21) 2 precursor - yeast (Saccharomycopsis fibuligera)  
C:Species: Saccharomycopsis fibuligera  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 24-Sep-1999  
C:Accession: B45956  
R:Mechida, M.; Ohtsuki, I.; Fukui, S.; Yamashita, I.  
Appl. Environ. Microbiol. 54, 3147-3155, 1988  
A:Title: Nucleotide sequences of Saccharomycopsis fibuligera genes for extracellular beta-glucosidase  
A:Reference number: A45956; MUID:89133518; PMID:3146949  
A:Accession: B45956  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-880 <MAC>  
A:Cross-references: GB:M22476; NID:g170809; PIDN:AAA34315.1; PID:g170810  
C:Superfamily: beta-glucosidase 1  
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 82.5%; Score 33; DB 2; Length 880;  
Best Local Similarity 71.4%; Pred. No. 67;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 WOSAFK 7  
11:1111  
DB 56 WOSAFK 62

## RESULT 4

S04795  
mobilization protein mobD - Escherichia coli plasmid ColK  
C:Species: Escherichia coli  
C:Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 18-Jun-1993  
C:Accession: S04795  
R:Boyd, A.C.; Archer, J.A.K.; Sherratt, D.J.  
Mol. Gen. Genet. 217, 488-498, 1989  
A:Title: Characterization of the ColE1 mobilization region and its protein products.  
A:Reference number: JQ0389; MUID:89364735; PMID:2671664  
A:Accession: S04795  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-77 <BOY>  
C:Genetics:  
A:Gene: mobD; mbxD  
A:Genome: plasmid

Query Match 80.0%; Score 32; DB 2; Length 77;  
Best Local Similarity 83.3%; Pred. No. 7.8;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WOSAFK 6  
11:1111  
DB 26 WOSAFK 31

## RESULT 5

C69451  
cationic amino acid transporter (cat-1) homolog - Archaeoglobus fulgidus  
C:Species: Archaeoglobus fulgidus  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 22-Oct-1999  
C:Accession: C69451  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson  
J.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
Nature 390, 364-370, 1997  
A:Authors: Uitterlinden, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.  
Smith, H.O.; Woese, C.R.; Venter, J.C.  
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae

A:Reference number: A69250; MUID:98049343; PMID:9389475  
A:Accession: C69451  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-736 <KLE>  
A:Cross-references: GB:AE000991; GB:AE000782; NID:g2689334; PIDN:ABB89635.1; PID:g264

Query Match 80.0%; Score 32; DB 2; Length 736;  
Best Local Similarity 83.3%; Pred. No. 89;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WOSAFK 6  
11:1111  
DB 186 WOSAFK 191

## RESULT 6

AD0159  
outer membrane usher protein PsaC precursor psac [imported] - Yersinia pestis (strain  
C:Species: Yersinia pestis  
C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Nov-2001  
C:Accession: AD0159  
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tiltball, R.W.; Holden, M.T.G.; Prentice, M.  
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.  
ll, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell  
Nature 413, 523-527, 2001  
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586360  
A:Accession: AD0159  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-837 <KUR>  
A:Cross-references: GB:AL590842; PIDN:CAC90135.1; PID:g15979355; GSPDB:GN00175  
C:Genetics:  
A:Gene: psac  
C:Superfamily: outer membrane usher protein fimD

Query Match 80.0%; Score 32; DB 2; Length 837;  
Best Local Similarity 83.3%; Pred. No. 1e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WOSAFK 6  
11:1111  
DB 220 WOSAFK 225

## RESULT 7

F86548  
polymorphic outer membrane protein E family [imported] - Chlamydia pneumoniae (st  
C:Species: Chlamydia pneumoniae  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2001  
C:Accession: F86548  
R:Shirai, M.; Hiraoka, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.;  
Nucleic Acids Res. 28, 2311-2314, 2000  
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.  
A:Reference number: AB6491; MUID:20330349; PMID:10871362  
A:Accession: F86548  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-938 <STO>  
A:Cross-references: GB:BA000008; NID:g8978836; PIDN:BAA98772.1; GSPDB:GN00142  
A:Experimental source: strain J138  
C:Genetics:  
A:Gene: pmp\_15

Query Match 80.0%; Score 32; DB 2; Length 938;  
Best Local Similarity 83.3%; Pred. No. 1.2e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WOSAFK 6  
11:1111  
DB 642 WOSAFK 647



RESULT 8  
H72074

polymorphic membrane protein E/F family CP0286 [imported] - Chlamydomonas pneumoniae (str. C:Species: Chlamydomonas pneumoniae, Chlamydia pneumoniae)

C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-May-2000

C:Accession: H72074; E81593

R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lamuel, C.; Fan, J.; Olinger, L.; Grimwood, J.; Nature Genet. 21, 385-389, 1999

A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.

A:Reference number: A72000; MUID:99206606; PMID:10192388

A:Accession: H72074

A:Molecule type: DNA

A:Residues: 1-938 <ARN>

A:Cross-references: GB:AE001631; GB:AE001363; NID:94376750; PIDN:AD18608.1; PID:9437675

A:Experimental source: strain CML029

R:Read, T.D.; Brunham, R.C.; Shen, C.; Gail, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000

A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.

A:Reference number: A81500; MUID:20150255; PMID:10684935

A:Accession: E81593

A:Molecule type: DNA

A:Residues: 1-938 <REA>

A:Cross-references: GB:AE002190; GB:AE002161; NID:97189209; PIDN:AAF38143.1; PID:9718921

A:Experimental source: strain AR39, HL cells

C:Genetics:

A:Gene: pmp\_15; CP0286

Query Match  
Best Local Similarity 83.3%; Score 32; DB 2; Length 938;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WSAFT 6  
Db 642 WSAFT 647

## RESULT 9

H82752  
hypothetical protein XF0867 [imported] - Xylella fastidiosa (strain 9a5c)

C:Species: Xylella fastidiosa

C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000

C:Accession: H82752

R:Anonymous. The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; MUID:20365717; PMID:10910347

A>Note: for a complete list of authors see reference number A59328 below

A:Accession: H82752

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-112 <SIM>

A:Cross-references: GB:AE003926; GB:AE003849; NID:99105771; PIDN:AAF83677.1; GSPDB:GN001

A:Experimental source: strain 9a5c

R:Simpton, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A. Biondo, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carrer, D.M.; Carrer, H. as-Neto, E.; Docena, C.; El-Dorri, H.; Facincent, A.P.; Ferreira, A.J.S.

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A59328

A:Contents: annotation

A:Gene: XF0867

Query Match 77.5%; Score 31; DB 2; Length 112;

Best Local Similarity 83.3%; Pred. No. 19;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WSAFT 6  
Db 95 WSAFT 100

## RESULT 10

H97480  
hypothetical protein AGR\_C\_1820 [imported] - Agrobacterium tumefaciens (strain C58, C

C:Species: Agrobacterium tumefaciens

C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 11-Jan-2002

C:Accession: H97480

R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Ouello, B.; Goldm

A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium

A:Reference number: A97359; PMID:11743194

A:Accession: H97480

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-289 <KUR>

A:Cross-references: GB:AE007869; PIDN:AAK86801.1; PID:915156007; GSPDB:GN00169

C:Genetics:

A:Gene: AGR\_C\_1820

Query Match 77.5%; Score 31; DB 2; Length 289;  
Best Local Similarity 83.3%; Pred. No. 53;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WSAFT 6  
Db 41 WSAFT 46

## RESULT 11

AH2698  
permease [imported] - Agrobacterium tumefaciens (strain C58, Dupont)

C:Species: Agrobacterium tumefaciens

C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002

C:Accession: AH2698

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Moo

A:Title: The genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; PMID:11743193

A:Accession: AH2698

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-289 <KUR>

A:Cross-references: GB:AE006688; PIDN:AA142006.1; PID:917739380; GSPDB:GN00186

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: AtU0992

A:Map position: circular chromosome

Query Match 77.5%; Score 31; DB 2; Length 289;  
Best Local Similarity 83.3%; Pred. No. 53;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WSAFT 6  
Db 41 WSAFT 46

## RESULT 12

I64069  
hypothetical protein HI0461 - Haemophilus influenzae (strain Rd KW20)

C:Species: Haemophilus influenzae  
 C>Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 29-Sep-1999  
 C:Accession: 164069  
 R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.  
 R:Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodex, A.; Kelley, J.M.; Weidman, J.  
 D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Furumann, J.L.; Geoghagen, N.S.M.  
 Science 265, 496-512, 1995  
 A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,  
 A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
 A:Reference number: A64000; MUID:95350630; PMID:7542800  
 A:Accession: 164069  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-291 <TIGR>  
 A:Cross-references: GB:U32728; GB:L42023; NID:g1573425; PIDN:AAC22119.1; PID:g1573435; T  
 C:Superfamily: conserved hypothetical protein HI0461

Query Match 77.5% Score 31; DB 2; Length 291;  
 Best Local Similarity 71.4% Pred. No. 53;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 MOSAFTR 7  
 ||| |||  
 Db 58 WHSGFTR 64

RESULT 13  
 G81185  
 hypothetical protein NMB0552 [imported] - Neisseria meningitidis (strain MC58 serogroup  
 C:Species: Neisseria meningitidis  
 C>Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
 C:Accession: G81185  
 R:Telletlin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.  
 Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;  
 Li, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scariato, V.; Maignani, V.; Pizze, M.  
 Science 287, 1809-1815, 2000  
 A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve  
 A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.  
 A:Reference number: A81000; MUID:20175755; PMID:10710307  
 A:Accession: G81185  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-357 <PEPT>  
 A:Cross-references: GB:AE002411; GB:AE002098; NID:g7225776; PIDN:AAF40981.1; PID:g722577  
 A:Experimental source: serogroup B, strain MC58  
 C:Genetics:

Query Match 77.5% Score 31; DB 2; Length 357;  
 Best Local Similarity 71.4% Pred. No. 67;  
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 MOSAFTR 7  
 ||| |||  
 Db 286 WHSGFTR 292

RESULT 14  
 F81916  
 probable integral membrane protein NMA0732 [imported] - Neisseria meningitidis (strain Z  
 C:Species: Neisseria meningitidis  
 C>Date: 03-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
 C:Accession: F81916  
 R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel  
 R; Holtz, S.; Jorgensen, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,  
 Nature 404, 502-506, 2000  
 A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.  
 A:Reference number: A81775; MUID:20222556; PMID:10761919  
 A:Accession: F81916  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-357 <PAR>  
 A:Cross-references: GB:AL162754; GB:AL157959; NID:g7379424; PIDN:CAB84016.1; PID:g737945

A:Experimental source: serogroup A, strain Z2491  
 C:Genetics:  
 A:Gene: NMA0732

Query Match 77.5% Score 31; DB 2; Length 357;  
 Best Local Similarity 71.4% Pred. No. 67;  
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 MOSAFTR 7  
 ||| |||  
 Db 286 WHSGFTR 292

RESULT 15  
 T32652  
 hypothetical protein F39C12.3 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 18-Feb-2000  
 C:Accession: T32652  
 R:Chisoe, S.; Sansone, J.  
 submitted to the EMBL Data Library, December 1997  
 A:Description: The sequence of C. elegans cosmid F39C12.  
 A:Reference number: Z21206  
 A:Accession: T32652  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-427 <CHI>  
 A:Cross-references: EMBL:AF039043; PIDN:AB94196.1; GSPDB:GN00028; CESP:F39C12.3  
 A:Experimental source: strain Bristol N2; clone F39C12  
 C:Genetics:  
 A:Gene: CESP:F39C12.3  
 A:Map position: X  
 A:Introns: 42/3; 104/3; 133/3; 164/3; 213/3; 276/3; 336/3

Query Match 77.5% Score 31; DB 2; Length 427;  
 Best Local Similarity 71.4% Pred. No. 81;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 MOSAFTR 7  
 ||| |||  
 Db 230 WQNALTR 236

Search completed: July 11, 2003, 12:00:34  
 Job time : 4.5164 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:43:41 ; Search time 0.669439 Seconds

(without alignments)  
433.698 Million cell updates/sec

Title: US-10-091-442-2  
Perfect score: 40  
Sequence: 1 MOSAFTRK 7

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt.40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40	100.0	380	1 SPB7_HUMAN	O75635 homo sapien
2	33	82.5	880	1 BGL2_SACFI	P22507 saccharomyc
3	32	80.0	369	1 H183_RHIL0	O98B00 rhizobium l
4	32	80.0	837	1 PSAC_YERPE	P31527 yersinia pe
5	32	80.0	837	1 PSAC_YERPS	O56983 yersinia ps
6	32	80.0	938	1 PM15_CHLPN	O92883 chlamydia p
7	32	80.0	959	1 G2D1_HUMAN	O9uh19 h general t
8	31	77.5	291	1 Y461_HAEIN	O57144 haemophilus
9	31	77.5	677	1 FUS2_YEAST	O05670 saccharomyc
10	31	77.5	722	1 GRE4_CHLPN	O92794 chlamydia p
11	31	77.5	978	1 PM14_CHLPN	O92895 chlamydia p
12	30	75.0	296	1 UPBS_MYCLE	P38119 mycobacteri
13	30	75.0	351	1 GSTB_BPT4	O94547 bacteriopho
14	30	75.0	385	1 PAR4_HUMAN	P04547 homo sapien
15	30	75.0	429	1 GLN4_PHAVU	P15102 phaseolus v
16	30	75.0	432	1 GLN2_DAUCA	O22506 daucus caro
17	30	75.0	494	1 TRPE_AQUAE	O66849 aquilex aeo
18	30	75.0	653	1 YDCP_ECOLI	P76104 escherichia
19	30	75.0	917	1 GCP3_DROME	O95yp8 drosophil
20	30	75.0	1210	1 YEH1_ECOLI	P33346 escherichia
21	29	72.5	137	1 V44_SOLIN	P33777 soleosopsis
22	29	72.5	225	1 Y991_COREL	P38118 corynebacte
23	29	72.5	259	1 Y114_STRFR	P20182 streptomyce
24	29	72.5	262	1 YAB6_MYCTU	O35434 mycobacteri
25	29	72.5	270	1 ALYS_BPRIT	O38135 bacteriopho
26	29	72.5	345	1 HEMS_YEREN	P31517 yersinia en
27	29	72.5	345	1 HMUS_YERPE	O56990 yersinia pe
28	29	72.5	351	1 UXUA_CLOAB	O971f4 clostridium
29	29	72.5	359	1 MOA1_MYCTU	O05786 mycobacteri
30	29	72.5	405	1 IFSP_MOUSE	P70458 mus musculu
31	29	72.5	450	1 DGT1_PASWU	O9cm11 pasteurella
32	29	72.5	481	1 GIGA_RHIL0	O985p2 rhizobium l
33	29	72.5	580	1 FZ0B_XENLA	O9w742 xenopus lae

34	29	72.5	586	1 FZ0A_XENLA	O9dab5 xenopus lae
35	29	72.5	601	1 YE06_SCHPO	O13803 schizosacch
36	29	72.5	639	1 V70K_PLR1	P17519 portato leaf
37	29	72.5	639	1 V70K_PLR1	P11622 portato leaf
38	29	72.5	639	1 V70K_PLR1	O9uxu2 pyrococcus
39	29	72.5	693	1 TGM3_HUMAN	O08188 homo sapien
40	29	72.5	714	1 GRE4_CHLMU	O9plu1 chlamydia m
41	29	72.5	787	1 YD5A_SCHPO	O10313 schizosacch
42	29	72.5	1124	1 PHY1_TORAC	P33530 nicotiana t
43	29	72.5	1935	1 YBBA_SCHPO	O60072 schizosacch
44	29	72.5	2376	1 YIM9_YEAST	P40468 saccharomyc
45	28	70.0	77	1 MBED_ECOLI	P13660 escherichia

## ALIGNMENTS

RESULT 1  
SPB7\_HUMAN STANDARD: PRT: 380 AA.  
ID SPB7\_HUMAN  
AC O75635:  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Megsin (TP55) (Serpin B7).  
GN SERPINB7.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_Taxid:9606;  
[1]  
RP MEDLINE-97326116; PubMed-9182567;  
RX Tsurumoto M., Tsurumoto K., Ishida N., Kurihara T., Iyasa F.,  
RA Yamashiro K., Rogi T., Kodama S., Katsunagi N., Adachi M.,  
RA Katayama T., Nakao M., Yamachi K., Hashino J., Haruyama M., Miura K.,  
RA Nakanishi T., Nakazato H., Teramura M., Mitsuuchi H., Yamaguchi N.,  
RT "Purification, cDNA cloning, and characterization of a new serpin with  
RT megakaryocyte maturation activity."  
RL J. Biol. Chem. 272:15373-15380(1997).  
[2]  
RN SEQUENCE FROM N.A.  
RP TISSUE-Mesangial cells;  
RC MEDLINE-98376492; PubMed-9710452;  
RX Miyata T., Nangaku M., Suzuki D., Inagi R., Uragami K., Sakai H.,  
RA Okubo K., Kurokawa K.;  
RT "A mesangium-predominant gene, megsin, is a new serpin upregulated in  
RT Iga nephropathy."  
RL J. Clin. Invest. 102:828-836(1998).  
CC - FUNCTION: Might function as an inhibitor of lys-specific  
CC proteases. Might influence the maturation of megakaryocytes via  
CC its action as a serpin.  
CC - SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC - TISSUE SPECIFICITY: Predominantly expressed in mesangial cells.  
CC - SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use. By non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC -----  
CC EMBL: D88575; BAA3123.1; -  
CC EMBL: AF027866; AAC64506.1; -  
CC HSSP: P05619; IHDE.  
CC Genew: HGNC:13902; SERPINB7.  
CC MIM: 603357; -  
CC InterPro: IPR000215; Serpin.  
CC Pfam: PF00079; serpin; 1.  
CC SMART: SM00093; SERPIN; 1.  
CC PROSITE: PS00284; SERPIN; 1.

KM Serpin; Serine protease inhibitor.  
 FT ACT\_SITE 347 348 REACTIVE BOND (BY SIMILARITY).  
 SQ SEQUENCE 380 AA; 42904 MW; 9A2CDB6C63CF605 CRC64;

Query Match 100.0%; Score 40; DB 1; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 0.36;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WOSAFTR 7  
 DB 181 WOSAFTR 187

RESULT 2  
 BG12\_SACFI STANDARD; PRT; 880 AA.  
 AC P22507;  
 DT 01-AUG-1991 (Rel. 19, Created)  
 DT 01-AUG-1991 (Rel. 19, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Beta-glucosidase 2 precursor (EC 3.2.1.21) (Gentloblase) (Cellloblase)  
 DE (beta-D-glucoside glucohydrolase).  
 GN BG12.  
 OS Saccharomycopsis fibuligera (Yeast)  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycopsidaceae; Saccharomycopsis.  
 OX NCBI\_TaxID=9444;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89133518; PubMed=1146949;  
 RA Machida M., Ohtsuki I., Fukui S., Yamashita I.;  
 RT "Nucleotide sequences of Saccharomycopsis fibuligera genes for  
 RT extracellular beta-glucosidases as expressed in Saccharomyces  
 RT cerevisiae";  
 RL Appl. Environ. Microbiol. 54:3147-3155(1988).  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-  
 CC glucose residues with release of beta-D-glucose.  
 CC -1- PATHWAY: Cellulose degradation.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 3 OF GLYCOSYL HYDROLASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M22476; AAA34315.1; -;  
 CC IPR: B45956; B45956.  
 DR InterPro: IPR002772; GH\_3C.  
 DR InterPro: IPR001764; GH\_3N.  
 DR Pfam: PF00933; Glyco\_hydro\_3\_1.  
 DR Pfam: PF01915; Glyco\_hydro\_3\_C\_1.  
 DR PRINTS: PR00133; GLHDLASE3.  
 DR PROSITE: PS00775; GLYCOSYL-HYDROL\_F3\_1.  
 DR PROSITE: PS00775; GLYCOSYL-HYDROL\_F3\_1.  
 KW Hydrolyase; Glycosidase; Cellulose degradation; Glycoprotein; Signal.  
 FT CHAIN 1 17  
 FT ACT\_SITE 299 299  
 FT CARBOHYD 24 24  
 FT CARBOHYD 77 77  
 FT CARBOHYD 271 271  
 FT CARBOHYD 336 336  
 FT CARBOHYD 343 343  
 FT CARBOHYD 376 376  
 FT CARBOHYD 548 548  
 FT CARBOHYD 589 589  
 FT CARBOHYD 712 712  
 FT CARBOHYD 743 743  
 FT CARBOHYD 794 794  
 SQ SEQUENCE 880 AA; 96799 MW; C95061283090814C CRC64;

Query Match 82.5%; Score 33; DB 1; Length 880;  
 Best Local Similarity 71.4%; Pred. No. 26;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 WOSAFTR 7  
 DB 56 WODAYTR 62

RESULT 3  
 H183\_RHIL0 STANDARD; PRT; 369 AA.  
 ID H183\_RHIL0  
 AC Q98B00;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Histidinol-phosphate aminotransferase 3 (EC 2.6.1.9) (Imidazole  
 DE acetol-phosphate transaminase 3).  
 GN HISC3 OR MLR5786.  
 OS Rhizobium loti (Mesorhizobium loti).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Mesorhizobium.  
 OX NCBI\_TaxID=381;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=MAF303099;  
 RX MEDLINE=11082930; PubMed=11214968;  
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
 RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kiyokawa C., Kohara M., Matsunoto M., Matsuno A.,  
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,  
 RA Takeuchi C., Yamada M., Tabata S.;  
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
 RT Mesorhizobium loti";  
 RL DNA Res. 7:331-338(2000).  
 CC -1- CATALYTIC ACTIVITY: L-histidinol-phosphate + 2-oxoglutarate = 3-  
 CC (imidazol-4-yl)-2-oxopropyl phosphate + L-glutamate.  
 CC -1- COFACTOR: Pyridoxal phosphate (By similarity).  
 CC -1- PATHWAY: Histidine biosynthesis; seventh step.  
 CC -1- SIMILARITY: BELONGS TO CLASS-II OF PYRIDOXAL-PHOSPHATE-DEPENDENT  
 CC AMINOTRANSFERASES. HISTIDINOL-PHOSPHATE AMINOTRANSFERASE  
 CC SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AP003007; BAB52172.1; -;  
 CC IPR: AP004839; AminoTransf1/2.  
 DR InterPro: IPR001917; NHTransf\_2.  
 DR Pfam: PF00155; aminoTransf\_2.  
 DR TIGRfam: TIGR01141; hisc3\_1.  
 DR PROSITE: PS00599; AA\_TRANSFER\_CLASS\_2; 1.  
 DR PROSITE: PS00599; AA\_TRANSFER\_CLASS\_2; 1.  
 KW Histidine biosynthesis; Transferase; Amino transferase;  
 KW Pyridoxal phosphate; Complete proteome.  
 FT BINDING 220 220  
 SQ SEQUENCE 369 AA; 40710 MW; DCB33E0507165570 CRC64;

Query Match 80.0%; Score 32; DB 1; Length 369;  
 Best Local Similarity 71.4%; Pred. No. 17;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 WOSAFTR 7  
 DB 315 WOSAFTR 321

RESULT 4  
 PSAC\_YERPE

ID PSAC\_YERPE STANDARD: PRT: 837 AA.  
AC P31537: 056980: 01-JUL-1993 (Rel. 26, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Outer membrane usher protein psac precursor.  
PSAC OR YP01305.  
GN Yersinia pestis.  
OS Yersinia pestis.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Yersinia.  
NCBI\_TaxID=632;  
OX NCBI\_TaxID=632;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN-EV76:  
RA Cherepanov P.A.;  
RL Submitted (May-1996) to the EMBL/GenBank/DBJ databases.  
RN (2)  
RP SEQUENCE FROM N.A.  
RC STRAIN-CO-92 / Blovat Orientalis;  
RX MEDLINE-21470413: PubMed-11586360;  
RA Parthall J., Wren B.W., Thomson N.R., Tibball R.W., Holden M.T.G.,  
RA Prentice M.B., Sepalnia M., James K.D., Churcher C., Mungall K.L.,  
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Farraga A.M.,  
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,  
RA Felkwell T., Hamlin N., Holroyd S., Jagsels K., Karlyshev A.V.,  
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,  
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;  
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";  
RL Nature 413:523-527(2001).  
RN (3)  
RP SEQUENCE OF 1-196 FROM N.A.  
RX MEDLINE-93302507: PubMed-8100346;  
RA Lindler L.E., Tall B.D.;  
RT "Yersinia pestis PH 6 antigen forms fimbriae and is induced by  
intercellular association with macrophages.";  
RL Microbiol. 8:311-324(1993).  
CC -1- FUNCTION: INVOLVED IN THE EXPORT AND ASSEMBLY OF PSAA (PH 6)  
CC FIMBRIAL SUBUNITS ACROSS THE OUTER MEMBRANE.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane  
CC (By similarity).  
CC -1- SIMILARITY: BELONGS TO THE FIMBRIAL EXPORT USHER FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: X97759: CAA66357.1: ALT\_INIT.  
DR EMBL: A414147: CAC90135.1: -;  
DR EMBL: M87713: AAA27664.1: ALT\_INIT.  
DR PIR: S27749: S27749.  
DR PIR: S32928: S32928.  
DR InterPro: IPR000015: Fimb\_usher.  
DR Pfam: PF00577: usher.1.  
DR PROSITE: PS01151: FIMBRIAL\_USHER.1.  
KW Outer membrane; Transmembrane; Fimbria; Transport; Signal;  
KW Complete proteome.  
FT SIGNAL 1 23 POTENTIAL.  
FT CHAIN 24 837 OUTER MEMBRANE USHER PROTEIN PSAC.  
FT CONFLICT 172 177 ANISER -> EISYKP (IN REF. 1 AND 3).  
FT CONFLICT 178 196 OPRONGYRSGTITLYPG -> SFVKMVDGTGVHNTSYIP  
V (IN REF. 3).  
FT CONFLICT 395 395 R -> A (IN REF. 1).  
FT CONFLICT 437 437 E -> K (IN REF. 1).  
FT CONFLICT 797 805 ETGGLLVQW -> KQSGYCM (IN REF. 1).  
SO SEQUENCE 837 AA: 91702 MW: D901C236C012B2A CRC64;

Query Match 80.0%; Score 32; DB 1; Length 837;  
Best Local Similarity 83.3%; Pred. NO. 40;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MOSAFT 6  
Db 220 MOSAFT 225

RESULT 5  
PSAC\_YERPS  
ID PSAC\_YERPS STANDARD: PRT: 837 AA.  
AC Q56963;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Outer membrane usher protein psac precursor.  
GN PSAC.  
OS Yersinia pseudotuberculosis.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Yersinia.  
NCBI\_TaxID=633;  
OX NCBI\_TaxID=633;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN-YP111;  
RX MEDLINE-96294755: PubMed-8698470;  
RA Yang Y., Merriam J.J., Mueller J., Isberg R.R.;  
RT "The psa locus is responsible for thermolabile binding of Yersinia  
pseudotuberculosis to cultured cells.";  
RL Infect. Immun. 64:2483-2489(1996).  
CC -1- FUNCTION: INVOLVED IN THE EXPORT AND ASSEMBLY OF PSAA (PH 6)  
CC FIMBRIAL SUBUNITS ACROSS THE OUTER MEMBRANE.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane  
CC (By similarity).  
CC -1- SIMILARITY: BELONGS TO THE FIMBRIAL EXPORT USHER FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: L76301: AAC37058.1: ALT\_INIT.  
DR InterPro: IPR000015: Fimb\_usher.  
DR Pfam: PF00577: usher.2.  
DR PROSITE: PS01151: FIMBRIAL\_USHER.1.  
KW Outer membrane; Transmembrane; Fimbria; Transport; Signal.  
FT SIGNAL 1 23 POTENTIAL.  
FT CHAIN 24 837 OUTER MEMBRANE USHER PROTEIN PSAC.  
SO SEQUENCE 837 AA: 91786 MW: 820F7896DE4BEAA1 CRC64;

Query Match 80.0%; Score 32; DB 1; Length 837;  
Best Local Similarity 83.3%; Pred. NO. 40;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MOSAFT 6  
Db 220 MOSAFT 225

RESULT 6  
PM15\_CHLPN  
ID PM15\_CHLPN STANDARD: PRT: 938 AA.  
AC Q92883;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Probable outer membrane protein pmp15 precursor (Polymorphic membrane  
protein 15).  
GN PMP15 OR CPN0466 OR CP0286.  
OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=83558;  
RN (1)

RP SEQUENCE FROM N.A.  
 RC STRAIN-CWL029;  
 RX MEDLINE-99206606; PubMed-10192388;  
 RA Kaiman S., Mitchell W., Marathe R., Iammell C., Fan J., Hyman R.W.,  
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis";  
 RL Nat. Genet. 21:385-389(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AR39;  
 RX MEDLINE-20150255; PubMed-10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,  
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
 RA Gyalin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,  
 RA Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis Mohn and Chlamydia  
 RT pneumoniae AR39";  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-J138;  
 RX MEDLINE-20330349; PubMed-10871362;  
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 RT from Japan and CML029 from USA";  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 CC -1- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)  
 CC (POTENTIAL)  
 CC -1- SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AEO01631; AAD18608.1; -;  
 DR EMBL: AEO02190; AAF38143.1; -;  
 DR EMBL: APO02546; BAA98672.1; -;  
 DR TIGR: CP0286; -;  
 DR InterPro: IPR003368; Chlamydia\_PMP.  
 DR InterPro: IPR003357; OMP.  
 DR Pfam: PF02385; OMP. 1.  
 DR Pfam: PF02415; DUF145. 1.  
 KM Outer membrane; Signal; Multigene family; Complete proteome.  
 FT SIGNAL 1 17 POTENTIAL.  
 FT CHAIN 18 938 PROBABLE OUTER MEMBRANE PROTEIN PMP15.  
 SQ SEQUENCE 938 AA; 102194 MW; A904AAB05B567455 CRC64;  
 Query Match 80.0%; Score 32; DB 1; Length 938;  
 Best Local Similarity 83.3%; Pred. NO. 45;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 WQSAFT 6  
 Db 642 WQSAFT 647  
 RESULT 7  
 ID G2D1\_HUMAN STANDARD: PRT; 959 AA.  
 AC 09UHL9: 095444; 09UHK8: 09U191; 08WVC4;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DE General transcription factor II-I repeat domain-containing protein 1  
 DE (GTF2I repeat domain containing protein 1) (Muscle TFI-I repeat  
 DE domain-containing protein 1) (General transcription factor I1) (Slow-  
 DE muscle-fiber enhancer binding protein) (USE BI binding protein)

DE (MUSTRD1/BEN) (Williams-Beuren syndrome chromosome region 11 protein).  
 GN GTF2IRD1 OR MUSTRD1 OR GTF3 OR CREM1 OR RBAP2 OR WBSCTRL1 OR WBSCTRL2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_Taxid:9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RC TISSUE=Muscle;  
 RX MEDLINE-98449952; PubMed-9774679;  
 RA O'Mahoney J.V., Guven K.L., Lin J., Joya J.E., Robinson C.S.,  
 RA Wade R.P., Hardeman E.C.;  
 RT "Identification of a novel slow-muscle-fiber enhancer binding protein,  
 RT MUSTRD1";  
 RL Mol. Cell. Biol. 18:6641-6652(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RC TISSUE=Colon carcinoma;  
 RX MEDLINE-99216421; PubMed-10198167;  
 RA Osborne L.R., Campbell T., Daradich A., Scherer S.W., Tsui L.-C.;  
 RT "Identification of a putative transcription factor gene (WBSCTRL1) that  
 RT is commonly deleted in Williams-Beuren syndrome";  
 RL Genomics 57:279-284(1999).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC TISSUE=Fetal brain;  
 RX MEDLINE-20037629; PubMed-10573005;  
 RA Tassabehji M., Carette M., Wilmet C., Donnai D., Read A.P.,  
 RA Metcalfe K.;  
 RT "A transcription factor involved in skeletal muscle gene expression is  
 RT deleted in patients with Williams syndrome";  
 RL Eur. J. Hum. Genet. 7:737-747(1999).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RX MEDLINE-20044629; PubMed-10575229;  
 RA Franke Y., Peoples R.J., Francke U.;  
 RT "Identification of GTF2IRD1, a putative transcription factor within  
 RT the Williams-Beuren syndrome deletion at 7q11.23";  
 RL Cytogenet. Cell Genet. 86:296-304(1999).  
 RN [5]  
 RP SEQUENCE FROM N.A. (ISOFORM 1), INTERACTION WITH RB1, AND MUTAGENESIS.  
 RC TISSUE=Cervical carcinoma, Placenta, and Fetal spleen;  
 RX MEDLINE-20115113; PubMed-10642537;  
 RA Yan X., Zhao X., Qian M., Guo N., Gong X., Zhu X.;  
 RT "Characterization and gene structure of a novel retinoblastoma-  
 RT protein-associated protein similar to the transcription regulator  
 RT TFI-I";  
 RL Biochem. J. 345:749-757(2000).  
 RN [6]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC TISSUE=Uterus;  
 RA Strausberg R.;  
 RT Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 RN [7]  
 RP FUNCTION.  
 RX MEDLINE-21332325; PubMed-11438732;  
 RA Tussie-Luna M.I., Bayarsaihan D., Ruddle F.H., Roy A.L.;  
 RT "Repression of TFI-I-dependent transcription by nuclear exclusion";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:7789-7794(2001).  
 CC -1- FUNCTION: May be a transcription regulator involved in cell-cycle  
 CC progression and skeletal muscle differentiation. May repress GTF2I  
 CC transcriptional functions, by preventing its nuclear residency, or  
 CC by inhibiting its transcriptional activation. May contribute to  
 CC slow-twitch fiber type specificity during myogenesis and in  
 CC regenerating muscles. Binds tropomyosin I slow-muscle fiber enhancer  
 CC (USE BI). Binds specifically and with high affinity to the EFG  
 CC sequences derived from the early enhancer of HOXC8 (By  
 CC similarity).  
 CC -1- SUBUNIT: Interacts with the retinoblastoma protein (RB1) via its  
 CC C-terminus.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms: 1 (shown here) and 2; are  
 CC produced by alternative splicing.

CC -1- TISSUE SPECIFICITY: Highly expressed in adult skeletal muscle,  
 CC heart, fibroblast, bone and fetal tissues. Expressed at lower  
 CC levels in all other tissues tested.  
 CC -1- DEVELOPMENTAL STAGE: Highly expressed in developing and  
 CC regenerating muscles, at the time of myofiber diversification.  
 CC -1- DOMAIN: The N-terminal half may have an activating activity.  
 CC -1- DISEASE: Haploinsufficiency of GTF2I1RD1 may be the cause of  
 CC certain cardiovascular and musculo-skeletal abnormalities observed  
 CC in Williams-Beuren syndrome (WBS), a rare developmental disorder.  
 CC It is a contiguous gene deletion syndrome involving genes from  
 CC chromosome band 7q11.23.  
 CC -1- SIMILARITY: BELONGS TO THE TFI-I FAMILY.  
 CC -1- SIMILARITY: CONTAINS 5 GTF2I-LIKE REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF118270; AAD14687.2; -  
 CC EMBL; AF104923; AAD27668.1; -  
 CC EMBL; AF151354; AAF19786.1; -  
 CC EMBL; AF156489; AAF17358.1; -  
 CC EMBL; AF089107; AAF21796.1; -  
 CC EMBL; BC018136; AAH18136.1; -  
 CC Gene; HGNC:4661; GTF2I1RD1.  
 CC MIM; 604318; -  
 CC MIM; 194050; -  
 CC InterPro: IPR004212; GTF2I.  
 CC Pfam: PF02946; GTF2I; 5.  
 CC Transcription regulation; Developmental protein; DNA-binding;  
 CC Nuclear protein; Repeat; Alternative splicing; Polymorphism;  
 CC Williams-Beuren syndrome.  
 CC KW REPEAT 128 203 GTF2I 1.  
 CC FT REPEAT 351 426 GTF2I 2.  
 CC FT REPEAT 565 640 GTF2I 3.  
 CC FT REPEAT 705 780 GTF2I 4.  
 CC FT REPEAT 802 877 GTF2I 5.  
 CC FT DOMAIN 898 905 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 CC FT DOMAIN 906 930 SER-RICH.  
 CC FT VARSPLIC 656 670 MISSING (IN ISOFORM 2).  
 CC FT VARIANT 652 652 M -> V (IN DBSNP:2301895).  
 CC FT MUTAGEN 898 959 MISSING: CYTOPLASMIC LOCALIZATION.  
 CC FT CONFLICT 111 111 G -> S (IN REF. 1 AND 2).  
 CC FT CONFLICT 378 378 R -> Q (IN REF. 5).  
 CC SQ SEQUENCE 959 AA; 106057 MW; 7DA3097879701540 CRC64;  
 CC -----  
 CC Query Match 80.0%; Score 32; DB 1; Length 959;  
 CC Best Local Similarity 71.4%; Pred. No. 47;  
 CC Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 CC -----  
 CC Oy 1 WQSAPTK 7  
 CC | 11111;  
 CC Db 20 WNSAFTR 26  
 CC -----  
 CC RESULT 8  
 CC Y461\_HAEIN  
 CC ID Y461\_HAEIN STANDARD: PRT: 291 AA.  
 CC AC 057144; 005022;  
 CC DT 01-NOV-1997 (Rel. 35, Created)  
 CC DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 CC DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 CC DE Protein H10461.  
 CC GN H10461.  
 CC OS Haemophilus influenzae.  
 CC OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 CC OC Haemophilus.  
 CC NCBI\_TaxID=727;  
 CC OX

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Rd / KW20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Felschmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fritchman J.L., Geoghegan N.S.M.,  
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RA "Whole-genome random sequencing and assembly of Haemophilus influenzae  
 RP Rd.";  
 RL Science 269:496-512(1995).  
 RN [2]  
 RP IDENTIFICATION BY MASS SPECTROMETRY.  
 RX MEDLINE=20137488; PubMed=10675023;  
 RA Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,  
 RA Gray C., Fountoulakis M.;  
 RT "Two-dimensional map of the proteome of Haemophilus influenzae";  
 RL Electrophoresis 21:411-429(2000).  
 CC -1- SIMILARITY: TO E.COLI YBJX AND WEAK, TO P.HAEMOLYTICA LAMP.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U32728; AAC22119.1; -  
 CC TIGR; H10461; -  
 CC Complete proteome.  
 CC SQ SEQUENCE 291 AA; 34846 MW; 699FDC4204A2E2A0 CRC64;  
 CC -----  
 CC Query Match 77.5%; Score 31; DB 1; Length 291;  
 CC Best Local Similarity 71.4%; Pred. No. 22;  
 CC Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 CC -----  
 CC Oy 1 WQSAPTK 7  
 CC | 11111;  
 CC Db 58 WQSLFTQ 64  
 CC -----  
 CC RESULT 9  
 CC FUS2\_YEAST  
 CC ID FUS2\_YEAST STANDARD: PRT: 677 AA.  
 CC AC 005670; 005023;  
 CC DT 01-NOV-1997 (Rel. 35, Created)  
 CC DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 CC DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 CC DE Nuclear fusion protein FUS2.  
 CC GN FUS2 OR YMR232W OR YMG959.14.  
 CC OS Saccharomyces cerevisiae (baker's yeast).  
 CC OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 CC NCBI\_TaxID=4932;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN-S288C;  
 CC RX MEDLINE=96032260; PubMed=7559752;  
 CC RA Elion E.A., Trueheart J., Fink G.R.;  
 CC "Fus2 localizes near the site of cell fusion and is required for both  
 CC cell fusion and nuclear alignment during zygote formation.";  
 CC J. Cell Biol. 130:1283-1296(1995).  
 CC RN [2]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN-S288C / AB972;  
 CC RA Skelton J., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;  
 CC Submitted (JUN-1995) to the EMBL/Genbank/DBJ databases.

```

CC -1- FUNCTION: PROMOTES CELL FUSION DURING ZYGOTE FORMATION.
CC -1- SUBCELLULAR LOCATION: LOCALIZES NEAR THE SITE OF CELL FUSION.
CC -1- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY (DH) DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X90752; CA62275.1; -
DR EMBL: Z49939; CA90203.1; -
DR SGD: S0004845; FUS2.
DR InterPro: IPR001331; GDS_CDC24.
DR InterPro: IPR000219; RHOGEP.
DR SMART: SM00325; RHOGEP; 1.
DR PROSITE: PS00741; DH_1; FALSE_NEG.
DR PROSITE: PS50010; DH_2; 1.
KW Fusion protein.
FT DOMAIN 112 326 DH.
FT CONFLICT 185 185 L->R (IN REF. 1).
FT CONFLICT 203 203 F->L (IN REF. 1).
FT CONFLICT 543 546 MISSING (IN REF. 1).
FT CONFLICT 589 677 HDTECILNLYIKVFLKYLILTIAGKKYLOKLENNMSLND
IATGOIKNLDILOCKSKSRMTKRMVTKMDPFPDPPGSGRY
VRKLFEI -> PRYRMVLYELVQSLKIFGNHCKMKKIPA
KRS (IN REF. 1).
SQ SEQUENCE 677 AA; 79036 MW; D7C89ED0738BA31F CRC64;

Query Match
Best Local Similarity 77.5%; Score 31; DB 1; Length 677;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 WOSAFK 7
DB 199 WOKIFK 205

RESULT 10
GREA_CHLPN STANDARD; PRT; 722 AA.
AC Q92764; Q9J0E9;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcription elongation factor grea (Transcript cleavage factor
DE grea).
GN GREA OR CPN0741 OR CP0004.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=99206606; PubMed=10192388;
RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).

RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Uitterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).

```

```

RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shira M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hatori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
CC -1- FUNCTION: NECESSARY FOR EFFICIENT RNA POLYMERASE TRANSCRIPTION
CC ELONGATION PAST TEMPLATE-ENCODING ARRESTING SITES. THE ARRESTING
CC SITES IN DNA HAVE THE PROPERTY OF TRAPPING A CERTAIN FRACTION
CC OF ELONGATING RNA POLYMERASES THAT PASS THROUGH, RESULTING IN
CC LOCKED TERNARY COMPLEXES. CLEAVAGE OF THE NASCENT TRANSCRIPT BY
CC CLEAVAGE FACTORS SUCH AS GREA OR GREB ALLOWS THE RESUMPTION OF
CC ELONGATION FROM THE NEW 3' TERMINUS. GREA RELEASES SEQUENCES OF
CC 2 TO 3 NUCLEOTIDES (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 1 GREA ASSOCIATED DOMAIN 1 (GRAD1).
CC -1- SIMILARITY: BELONGS TO THE GREA/GREB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE001656; AAD18880.1; -
DR EMBL: AE002164; AAF37801.1; -
DR EMBL: AP002547; BAA98948.1; -
DR TIGR: CP0004; -
DR InterPro: IPR001437; Grea_Greb.
DR Pfam: PF03449; Grea_Greb_N; 1.
DR Prodom: PD004918; Grea_Greb; 1.
DR PROSITE: PS00830; GREAB_2; 1.
DR PROSITE: PS00828; GREAB_1; 1.
DR PROSITE: PS00830; GREAB_2; 1.
KW Transcription regulation; DNA-binding; Coiled coil; Complete proteome.
FT DOMAIN 1 504 GRAD1.
FT DOMAIN 565 722 GREB.
FT DOMAIN 606 644 COILED COIL (POTENTIAL).
SQ SEQUENCE 722 AA; 82642 MW; E280EDB124D5E86F CRC64;

Query Match
Best Local Similarity 77.5%; Score 31; DB 1; Length 722;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 WOSAFK 7
DB 250 WOSAKTK 256

RESULT 11
PML4_CHLPN STANDARD; PRT; 978 AA.
AC Q92895; Q9RB63;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable outer membrane protein pml4 precursor (Polymorphic membrane
DE protein 14).
GN PMP14 OR CPN0454 OR CP0298.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=99206606; PubMed=10192388;
RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).

```



```

RN [12]
RP SEQUENCE FROM N.A.
RC STRAIN-AR39:
RX MEDLINE-20150255: PubMed-10684935:
RA Read T.D., Brunham R.C., Shen C., Gail S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Uterback T., Berry K., Baas S.,
RA Linher K., Weidman J., Khouli H., Craven B., Bowman C., Dodson R.,
RA Gwyn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.:
RT "Genome sequences of Chlamydia trachomatis M09 and Chlamydia
RT pneumoniae A39."
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [13]
RP SEQUENCE FROM N.A.
RC STRAIN-J138:
RX MEDLINE-20330349: PubMed-10871362:
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.:
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CML029 from USA."
RL Nucleic Acids Res. 28:2311-2314(2000).
CC -1- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)
CC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE001629; AAD18596.1; -
DR EMBL: AE002191; AAF38155.1; -
DR EMBL: AP002546; BAA98661.1; -
DR PHC1: ZDPAGE; Q92895; -
DR TIGR: CP0298; -
DR InterPro: IPR003368; Chlamydia_PMP.
DR InterPro: IPR003357; OMP.
DR Pfam: PF02385; OMP; 1.
DR Pfam: PF02415; DUF145; 1.
RW Outer membrane; Signal; Multigene family; Complete proteome.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 978 PROBABLE OUTER MEMBRANE PROTEIN PMP14.
FT CONFLICT 379 379 N->D (IN REF. 1).
SQ SEQUENCE 978 AA; 103654 MW; 6EE142999D3019 CRC64;
Query Match 77.5%; Score 31; DB 1; Length 978;
Best Local Similarity 71.4%; Pred. NO. 77;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 1 WQSAFTK 7
DB 854 WQSKFTE 860
RESULT 12
UPPS_MYCLE STANDARD: PRT: 296 AA.
ID UPPS_MYCLE
AC P38119;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Undecaprenyl pyrophosphate synthetase (EC 2.5.1.33) (UPP synthetase)
DE (Dl-trans-poly-cis-decaprenyltransferase) (Undecaprenyl diphosphate
DE synthase) (UDS).
GN UPPS OR ML0634 OR B1937_F2_65.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RA Smith D.R., Robison K.:
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RN [12]
RP SEQUENCE FROM N.A.
RC STRAIN-TN:
RX MEDLINE-21128732: PubMed-11234002:
RA Cole S.T., Eigleiner K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garner T., Churcher C., Harris D.,
RA Wungail K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltham T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagers K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajadream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.:
RT "Massive gene decay in the leprosy bacillus."
RL Nature 409:1007-1011(2001).
CC -1- FUNCTION: GENERATES UNDECAPRENYL PYROPHOSPHATE (UPP) FROM
CC ISOPRENYL PYROPHOSPHATE (IPP). UPP IS THE PRECURSOR OF THE
CC CARRIER LIPID FOR PEPTIDOGLYCAN SYNTHESIS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: dl-trans-poly-cis-decaprenyl diphosphate +
CC isopentenyl diphosphate -> diphosphate + dl-trans-poly-cis-
CC undecaprenyl diphosphate.
CC -1- SIMILARITY: BELONGS TO THE UPP SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U00016; AA17169.1; -
DR EMBL: AL583919; CAC30142.1; -
DR Leprosin: ML0634; -
DR InterPro: IPR001441; UPP_synth.
DR Pfam: PF01255; UPP_synthetase; 1.
DR ProDom: PD003461; UPP_synth; 1.
DR TIGRFAMs: TIGR00055; upps; 1.
DR PROSITE: PS01066; UPP_SYNTHETASE; 1.
RW Transferase; Cell division; Cell wall; Peptidoglycan synthesis;
KW Complete proteome.
SQ SEQUENCE 296 AA; 33890 MW; FF4D28BA49E76C545 CRC64;
Query Match 75.0%; Score 30; DB 1; Length 296;
Best Local Similarity 57.1%; Pred. NO. 36;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
OY 1 WQSAFTK 7
DB 257 WQAAVTE 263
RESULT 13
GSTR_BPT4 STANDARD: PRT: 351 AA.
ID GSTR_BPT4
AC P04547; Q38417;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE DNA beta-glucosyltransferase (EC 2.4.1.27) (BGT).
GN BGT OR BETA-GT.
OS Bacteriophage T4.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC T4-like viruses.
OX NCBI_TaxID=10665;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-86067181: PubMed-2999696;
RA Tomaszewski J., Gram H., Grubb J.W., Rueger W.:
RT "T4-induced alpha- and beta-glucosyltransferase: cloning of the genes

```

RT and a comparison of their products based on sequencing data.";  
 RL Nucleic Acids Res. 13:7551-7568(1985).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Kutter E., Arita F., Kunisawa T., Tsugita A., Mosig G.,  
 RA Mesyanzhinov V., Ruger W., Stidham T., Thomas E.,  
 RT "Bacteriophage T4 genome analysis."  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 1-88 FROM N.A.  
 RX MEDLINE-88169543; PubMed-2832395;  
 RA Thylen C.,  
 RT "Expression and DNA sequence of the cloned bacteriophage T4 dCMP  
 RT hydroxymethylase gene."  
 RL J. Bacteriol. 170:1994-1998(1988).  
 RN [4]  
 RP SEQUENCE OF 1-10 FROM N.A.  
 RX MEDLINE-87231080; PubMed-3295783;  
 RA Lamm N., Tomaszewski J., Rueger W.,  
 RT "Nucleotide sequence of the deoxycytidylate hydroxymethylase gene of  
 RT bacteriophage T4 (942) and the homology of its gene product with  
 RT thymidylate synthase of E. coli."  
 RL Nucleic Acids Res. 15:3920-3920(1987).  
 RN [5]  
 RP SEQUENCE OF 239-351 FROM N.A.  
 RX MEDLINE-92335355; PubMed-1631169;  
 RA Sharma M., Ellis R.L., Hinton D.M.,  
 RT "Identification of a family of bacteriophage T4 genes encoding  
 RT proteins similar to those present in group I introns of fungi and  
 RT phage."  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:6658-6662(1992).  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).  
 RX MEDLINE-94341252; PubMed-8062817;  
 RA Veilink A., Rueger W., Driessen H.P.C., Freemont P.S.,  
 RT "Crystal structure of the DNA modifying enzyme beta-  
 RT glucosyltransferase in the presence and absence of the substrate  
 RT uridine diphosphoglucose."  
 RL EMBO J. 13:3413-3422(1994).  
 RN [7]  
 RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).  
 RX MEDLINE-99428644; PubMed-10497034;  
 RA Moreta S., Imberly A., Aschke-Sonnenborn U., Ruger W., Freemont P.S.,  
 RT "T4 phage beta-glucosyltransferase: substrate binding and proposed  
 RT catalytic mechanism."  
 RL J. Mol. Biol. 292:717-730(1999).  
 CC -1- FUNCTION: CATALYZES THE TRANSFER OF GLUCOSE (GLC) FROM URIDINE  
 CC DIPHOSPHOGLUCOSE (UDP-GLC) TO 5-HYDROXYMETHYLCYTOSINE (5-HMC) IN  
 CC DOUBLE-STRANDED DNA.  
 CC -1- CATALYTIC ACTIVITY: Transfers a beta-D-glucosyl residue from UDP-  
 CC glucose to an hydroxymethylcytosine residue in DNA.  
 CC -1- PATHWAY: INVOLVED IN THE DNA MODIFICATION PROCESS TO PROTECTS  
 CC THE PHAGE GENOME AGAINST ITS OWN NUCLEASES AND THE HOST  
 CC RESTRICTION ENDONUCLEASE SYSTEM.  
 CC -1- SUBUNIT: MONOMER.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X03139; CAA26908.1; -  
 CC EMBL: AF158101; AAD42545.1; -  
 CC EMBL: M22767; AAA88469.1; -  
 CC EMBL: M69268; AAA32544.1; -  
 CC EMBL: Y00148; CAA68343.1; -  
 CC PIR: A00576; XUBP84.  
 CC PDB: 1BGT; 30-SEP-94.  
 CC PDB: 1BGU; 30-SEP-94.  
 CC PDB: 2BGT; 09-DEC-95.

DR PDB: 2BGT; 09-DEC-95.  
 DR PDB: 1OKJ; 28-JUL-99.  
 DR PDB: 1C3J; 09-AUG-99.  
 KW Transferase; Glycosyltransferase; 3D-structure.  
 FT CONFLICT 85 KFWA -> NLMQ (IN REF. 3).  
 SQ SEQUENCE 351 AA; 40666 MW; D1F42B5FE6CB9D61 CRC64;  
 Query Match 75.0%; Score 30; DB 1; Length 351;  
 Best Local Similarity 71.4%; Pred. No. 43;  
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 WSAFTK 7  
 DDb 341 WDAFFK 347  
 RESULT 14  
 PAR4\_HUMAN STANDARD; PRT; 385 AA.  
 AC 096R10; 076067;  
 DT 15-JUN-2002 (Ref. 41, Created)  
 DT 15-JUN-2002 (Ref. 41, Last sequence update)  
 DT 15-JUN-2002 (Ref. 41, Last annotation update)  
 DE Proteinase activated receptor 4 precursor (PAR-4) (Thrombin receptor-  
 DE like 3) (Coagulation factor II receptor-like 3).  
 GN F2RL3 OR PAR4.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-98389762; PubMed-9722561;  
 RA Kahn M.L., Hammes S.R., Botka C., Coughlin S.R.,  
 RT "Gene and locus structure and chromosomal localization of the  
 RT protease-activated receptor gene family."  
 RL J. Biol. Chem. 273:23290-23296(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-98379991; PubMed-9716134;  
 RA Kahn M.L., Zheng Y.-W., Huang W., Bigornia V., Zeng D., Moff S.,  
 RA Farese R.V., Jr., Tam C., Coughlin S.R.,  
 RT "A dual thrombin receptor system for platelet activation."  
 RL Nature 394:690-694(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A., AND MUTAGENESIS OF ARG-47 AND ARG-68.  
 RC TISSUE=lymphoma;  
 RX MEDLINE-98283984; PubMed-9618465;  
 RA Xu W.-F., Andersen H., Whitmore T.E., Presnell S.R., Yee D.P.,  
 RA Ching A., Gilbert T., Davie E.W., Foster D.C.,  
 RT "Cloning and characterization of human protease-activated receptor  
 RT 4."  
 RL Proc. Natl. Acad. Sci. U.S.A. 95:6642-6646(1998).  
 RN [4]  
 RP SEQUENCE FROM N.A., AND VARIANTS VAL-296 AND LEU-310.  
 RA Rieder M.J., Carrington D.P., Chung M.-W., Lee K.L., Peol C.L., Yi Q.,  
 RA Nickerson D.A.,  
 RT Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP FUNCTION  
 RX MEDLINE-99178892; PubMed-10079109;  
 RA Kahn M.L., Nakanishi-Matsui M., Shapiro M.J., Ishihara H.,  
 RA Coughlin S.R.,  
 RT "Protease-activated receptors 1 and 4 mediate activation of human  
 RT platelets by thrombin."  
 RL J. Clin. Invest. 103:879-887(1999).  
 CC -1- FUNCTION: Receptor for activated thrombin or trypsin coupled to G  
 CC proteins that stimulate phosphoinositide hydrolysis. May play a  
 CC role in platelets activation.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- TISSUE SPECIFICITY: Widely expressed, with highest levels in lung,  
 CC pancreas, thyroid, testis and small intestine. Not expressed in  
 CC brain, kidney, spinal cord and peripheral blood leukocytes. Also

```

DE 15-JUN-2002 (Ref. 41, Last annotation update)
DE Glutamine synthetase leaf isozyme, chloroplast precursor (EC 6.3.1.2)
OS (isozyme delta) (Glutamate--ammonia ligase).
OS Phaseolus vulgaris (Kidney bean) (French bean).
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;
OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae;
OC eurosids I; Fabales: Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.
OX NCBI_TaxID=3885;
RN [1]
RP SEQUENCE FROM N.A.
RC SPRAIN-cv. Tendergreen; TISSUE=Leaf;
RA Lightfoot D.A., Green N.K., Cullimore J.V.;
RT "The chloroplast-located glutamine synthetase of Phaseolus vulgaris
RT L.: nucleotide sequence, expression in different organs and uptake
RT into isolated chloroplasts."
RT Plant Mol. Biol. 11:191-202(1988).
RL -1- FUNCTION: THE LIGHT-MODULATED CHLOROPLAST ENZYME, ENCODED BY A
CC NUCLEAR GENE AND EXPRESSED PRIMARILY IN LEAVES, IS RESPONSIBLE FOR
CC THE REASSIMILATION OF THE AMMONIA GENERATED BY PHOTORESPIRATION.
CC -1- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +
CC L-glutamine.
CC -1- SUBUNIT: HOMOOCTAMER.
CC -1- SUBCELLULAR LOCATION: Chloroplast.
CC -1- MISCELLANEOUS: THERE ARE AT LEAST FOUR ISOZYMES OF THIS ENZYME IN
CC P. VULGARIS.
CC -1- MISCELLANEOUS: IRREVERSIBLY INHIBITED BY THE HERBICIDE
CC L-PHOSPHATONICRIN (PPT).
CC -1- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; X12738; CAA31234.1; -.
CC DR PIR; S04031; AUFBOD.
CC DR InterPro; IPR001691; GLN_synth.
CC DR InterPro; IPR001637; GLN_adenyln.
CC DR Pfam; PF00120; gln-synt; 1.
CC DR PROSITE; PS00180; GLNA_1; 1.
CC DR PROSITE; PS00181; GLNA_ATP; 1.
CC KW Nitrogen fixation; Ligase; Multigene family; Chloroplast;
CC KM Transist peptide.
CC FT TRANSIT 1 57 CHLOROPLAST.
CC FT CHAIN 58 429 GLUTAMINE SYNTHETASE LEAF ISOZYME.
CC SQ SEQUENCE 429 AA; 47246 MW; 0CA35624B118A78 CRC64;
CC
CC Query Match 75.0%; Score 30; DB 1; Length 429;
CC Best Local Similarity 71.4%; Pred. No. 53;
CC Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
CC
CC OY 1 WOSAF7K 7
CC II III
CC Db 11 WOMBFTK 17

```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:44:31 ; Search time 3.05613 Seconds  
(Without alignments)  
471.946 Million cell updates/sec

Title: US-10-091-442-2

Perfect score: 40

Sequence: 1 MOSAFPK 7

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: SP archaea:\*  
2: SP bacteria:\*  
3: SP fungi:\*  
4: SP human:\*  
5: SP invertebrate:\*  
6: SP mammal:\*  
7: SP mhc:\*  
8: SP organelle:\*  
9: SP phage:\*  
10: SP plant:\*  
11: SP rodent:\*  
12: SP virus:\*  
13: SP vertebrate:\*  
14: SP unclassified:\*  
15: SP viirus:\*  
16: SP bacteriophage:\*  
17: SP archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query	Length	DB ID	Description
1	36	90.0	11	Q9D695	09D695 mus musculus
2	36	90.0	11	Q9D695	09D695 mus musculus
3	36	90.0	11	Q9D695	09D695 mus musculus
4	34	85.0	313	Q9REB9	09REB9 bradyrhizob
5	34	85.0	1800	Q94882	094882 dirosophila
6	33	82.5	302	10 Q9LUV1	09LUV1 arabidopsis
7	33	82.5	397	16 Q9KFE1	09KFE1 bacillus ha
8	33	82.5	865	3 Q43022	043022 schizosacch
9	32	80.0	77	2 Q52995	052995 escherichia
10	32	80.0	161	16 Q92RF9	092RF9 rhizobium m
11	32	80.0	385	16 Q9B000	09B000 rhizobium l
12	32	80.0	386	3 Q9UV62	09UV62 aspergillus
13	32	80.0	386	3 Q9P900	09P900 aspergillus
14	32	80.0	435	10 Q9FFV0	09FFV0 arabidopsis
15	32	80.0	473	5 Q9GRW8	09GRW8 dirosophila
16	32	80.0	712	2 Q87537	087537 shewanella

17	32	80.0	736	17 Q28661	028661 archaeoglob
18	32	80.0	750	16 Q8XJ85	08XJ85 clostridium
19	32	80.0	1072	5 Q9KRV1	09KRV1 dirosophila
20	32	80.0	1080	5 Q9Y167	09Y167 dirosophila
21	31	77.5	112	16 Q9PE11	09PE11 xyella fas
22	31	77.5	187	10 Q9LUX6	09LUX6 arabidopsis
23	31	77.5	243	16 Q8ZP26	08ZP26 salmonella
24	31	77.5	275	10 Q9XHV1	09XHV1 oryza sativa
25	31	77.5	289	16 Q8UCP3	08UCP3 agrobacteri
26	31	77.5	357	16 Q9KON5	09KON5 neisseria m
27	31	77.5	357	16 Q9JVR3	09JVR3 neisseria m
28	31	77.5	427	5 Q44582	044582 caenorhabdit
29	31	77.5	451	5 Q95X63	095X63 caenorhabdit
30	31	77.5	514	17 Q9V052	09V052 pyrococcus
31	31	77.5	554	17 Q9HFE0	09HFE0 thermoplasma
32	31	77.5	558	17 Q978S9	0978S9 thermoplasma
33	31	77.5	570	5 Q9U4U9	09U4U9 haemoglobin
34	31	77.5	590	16 Q92X68	092X68 rhizobium m
35	31	77.5	666	16 Q9A107	09A107 streptococc
36	31	77.5	729	2 Q9EY54	09EY54 vibrio sp.
37	31	77.5	1108	16 Q8Y6V7	08Y6V7 listeria mo
38	31	77.5	1145	10 Q04146	004146 acetabulari
39	30	75.0	128	16 Q971M8	0971M8 clostridium
40	30	75.0	156	2 Q9F127	09F127 bacillus ci
41	30	75.0	163	3 Q8X1D5	08X1D5 ustilago ma
42	30	75.0	165	16 Q05500	005500 bacillus su
43	30	75.0	226	2 Q9ZB14	09ZB14 leuconostoc
44	30	75.0	228	8 Q9XUG8	09XUG8 leuconostoc
45	30	75.0	241	16 Q98BP8	098BP8 rhizobium l

## ALIGNMENTS

## RESULT 1

Q9D695 ID Q9D695 PRELIMINARY; PRT; 380 AA.  
AC Q9D695: 01-JUN-2001 (TREMUR1, 17, Created)  
DT 01-JUN-2001 (TREMUR1, 17, Last sequence update)  
DT 01-JUN-2002 (TREMUR1, 21, Last annotation update)  
DE 4631416M05R1k protein (Megsin).  
GN SERPINB7 OR 4631416M05R1k.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPRAIN-C57BL/6J; TISSUE=SKIN;  
RA MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishi Y.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,  
RA Saito T., Okazaki Y., Gotohori T., Bono H., Resakawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
RA Schirini L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonard M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombert P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilmberg L.,  
RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kotsuki S.,  
RA Hayashizaki Y.,  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
RN [2]  
RP SEQUENCE FROM N.A.

```

RX MEDLINE-21368006; PubMed-11473647;
RA Nangaku M., Miyata T., Suzuki D., Umezono T., Hashimoto T., Wada T.,
RA Yagi M., Nagano N., Inagi R., Kurokawa K.;
RT "Cloning of rodent megin revealed its up-regulation in
RT mesangiolipid proliferative nephritis.";
RL Kidney Int. 60:641-652(2001).
DR EMBL: AK014524; BAB29410.1;
DR EMBL: AF105328; AAL16768.1;
DR HSSP: P05121; 1A7C.
DR MGD: MGI:2151053; Serpinb7.
DR InterPro: IPR000215; Serpin.
DR Pfam: PF00079; serpin_1.
DR SMART: SM00093; SERPIN_1.
DR PROSITE: PS00284; SERPIN; UNKNOWN_1.
SQ SEQUENCE 380 AA; 43050 MW; C9240272BCE9CFA CRC64;

Query Match          90.0%; Score 36; DB 11; Length 380;
Best Local Similarity 85.7%; Pred. No. 21;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WQSAFTK 7
DB 181 WQSAFTK 187

RESULT 2
ID Q920J5 PRELIMINARY; PRT; 380 AA.
AC Q920J5;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE Megsin.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID-10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-21368006; PubMed-11473647;
RA Nangaku M., Miyata T., Suzuki D., Umezono T., Hashimoto T., Wada T.,
RA Yagi M., Nagano N., Inagi R., Kurokawa K.;
RT "Cloning of rodent megin revealed its up-regulation in
RT mesangiolipid proliferative nephritis.";
RL Kidney Int. 60:641-652(2001).
CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.
DR EMBL: AF105329; AAL16769.1;
DR InterPro: IPR000215; Serpin.
DR Pfam: PF00079; serpin_1.
DR PROSITE: PS00284; SERPIN; UNKNOWN_1.
KW Serpin.
SQ SEQUENCE 380 AA; 42821 MW; D8076CABEEZC2FBC CRC64;

Query Match          90.0%; Score 36; DB 11; Length 380;
Best Local Similarity 85.7%; Pred. No. 21;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WQSAFTK 7
DB 181 WQSAFTK 187

RESULT 3
ID Q9REE9 PRELIMINARY; PRT; 313 AA.
AC Q9REE9;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE Hypothetical 33.2 kDa protein.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Bradyrhizobium group; Bradyrhizobium.

```

```

OX NCBI_TaxID-375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-110SPC4;
RX MEDLINE-97261868; PubMed-9108282;
RA Minder A.C., Narberhaus F., Babs M., Hennecke H., Fischer H.M.;
RT "The dnaK operon belongs to the sigma32-dependent class of heat shock
RT genes in Bradyrhizobium japonicum.";
RL Mol. Gen. Genet. 254:195-206(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-110SPC4;
RX MEDLINE-20082837; PubMed-10613857;
RA Minder A.C., Fischer H.M., Hennecke H., Narberhaus F.;
RT "Role of HrcA and CIRCE in the heat shock regulatory network of
RT Bradyrhizobium japonicum.";
RL J. Bacteriol. 182:14-22(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-110SPC4;
RX MEDLINE-21150465; PubMed-11251836;
RA Minder A.C., de Rudder K.E.E., Narberhaus F., Fischer H.M.,
RA Hennecke H., Geiger O.;
RT "Phosphatidylcholine levels in Bradyrhizobium japonicum membranes are
RT critical for an efficient symbiosis with the soybean host plant.";
RL Mol. Microbiol. 39:1186-1198(2001).
DR EMBL: Y09633; CAB60667.1;
KW Hypothetical protein.
SQ SEQUENCE 313 AA; 33241 MW; 2F530355FCEB1B59 CRC64;

Query Match          85.0%; Score 34; DB 2; Length 313;
Best Local Similarity 71.4%; Pred. No. 45;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WQSAFTK 7
DB 295 WQSAFTK 301

RESULT 4
ID Q94882 PRELIMINARY; PRT; 1800 AA.
AC Q94882; 001381;
DT 01-FEB-1997 (TREMblrel. 02, Created)
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE DILUTE CLASS UNCONVENTIONAL MYOSIN.
DE DIDUM OR CG2146.
GN Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID-7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Maciver B.;
RL Thesis (1997), University of Edinburgh, Edinburgh, UK.
DR EMBL: Y08160; CAA69352.1;
DR HSSP: P10587; 1BR2.
DR FlyBase: FBgn0015933; didum.
DR InterPro: IPR002710; didum.
DR InterPro: IPR000048; IQ_region.
DR InterPro: IPR001609; myosin_head.
DR Pfam: PF01843; DIL; 1.
DR Pfam: PF00612; IQ; 6.
DR Pfam: PF00063; myosin_head; 1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR PRODOM: PD000376; myosin_head; 1.
DR PRODOM: PD000376; DIL; 1.
DR SMART: SM00015; IQ; 6.
DR SMART: SM00242; MYSC; 1.
DR PROSITE: PS50096; IQ; 1.
KW Myosin.

```

SQ SEQUENCE 1800 AA; 208948 MW; E88B0156FAAC57F1 CRC64;

Query Match Best Local Similarity 85.0%; Score 34; DB 5; Length 1800;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 WQSAFTK 7  
11:1111  
DB 649 WETAFTK 655

RESULT 5

O9LUY1 PRELIMINARY: PRT: 302 AA.

AC O9LUY1:

DT 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE C1a-prenyltransferase homolog-like, undecaprenyl phosphate synthetase-like.

OS Arabidopsis thaliana (mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.

OX NCBI\_TaxID=3702;

OC NCBI\_TaxID=3702;

RE SEQUENCE FROM N.A.

RC STRAIN-COLUMBIA;

RA MEDLINE=20181125; PubMed=10718197;

RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,

RA Tabata S.; "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence

RT features of the regions of 3,076,755 bp covered by sixty pl and YAC

RT clones.";

RL DNA Res. 7:31-63(2000).

DR EMBL; AB020755; BAA97348.1; "

DR InterPro: IPR001993; Mitoch.carrier.

DR InterPro: IPR001441; UPP\_synth.

DR Pfam: PF01255; UPP\_synthetase; 1.

DR Pfam: PF003461; UPP\_synth; 1.

DR Tricfams: TIGR00055; UPPS; 1.

DR PROSITE: PS00215; MITOCH\_CARRIER; UNKNOWN\_1.

DR PROSITE: PS01066; UPP\_SYNTHETASE; UNKNOWN\_1.

KW Transferase.

SQ SEQUENCE 302 AA; 34893 MW; C35020DDF493ABE CRC64;

Query Match Best Local Similarity 82.5%; Score 33; DB 10; Length 302;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 WQSAFTK 7  
11:1111  
DB 262 WQSAFTE 268

RESULT 6

O9KFB1 PRELIMINARY: PRT: 397 AA.

AC O9KFB1:

DT 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE Chromate transporter.

GN CHRA OR BH0574.

OS Bacillus halodurans.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;

OC Bacillaceae; Bacillus.

OX NCBI\_TaxID=86665;

RE SEQUENCE FROM N.A.

RC STRAIN-C-125 / JCM 9153;

RA MEDLINE=20512582; PubMed=11058132;

RA Takemi H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,

RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,

RA Horikoshi K.;

RT "Complete genome sequence of the alkaliphilic bacterium Bacillus

RT halodurans and genomic sequence comparison with Bacillus subtilis."

RL Nucleic Acids Res. 28:4317-4331(2000).

DR EMBL; AP001509; BAB04293.1; "

DR InterPro: IPR003370; Chromate\_transp.

DR Pfam: PF02417; Chromate\_transp; 2.

KW Complete proteome.

SQ SEQUENCE 397 AA; 42918 MW; 1E11CE27DB61ED56 CRC64;

Query Match Best Local Similarity 82.5%; Score 33; DB 16; Length 397;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 WQSAFTK 7  
11:1111  
DB 160 WQSAFTQ 166

RESULT 7

O43022 PRELIMINARY: PRT: 865 AA.

AC O43022:

DT 01-JUN-1998 (TREMBlrel. 06, Created)

DT 01-JUN-1998 (TREMBlrel. 17, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 19, Last annotation update)

DE Hypothetical 99.2 kDa protein C354.08C in chromosome II.

GN SPBC354.08C.

OS Schizosaccharomyces pombe (Fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomycetes.

OX NCBI\_TaxID=4896;

OC NCBI\_TaxID=4896;

RE SEQUENCE FROM N.A.

RC STRAIN-972;

RA Wood V., Rajandream M.A., Barrell B.G., Hiltbert H., Duesterhoeft A.;

RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).

CC -1- SIMILARITY: TO A. THALIANA ML23.19.

DR EMBL; AL022071; CAAL17808.1; "

DR InterPro: IPR003864; DUF221.

DR Pfam: PF02714; DUF221; 1.

KW Hypothetical protein; Transmembrane.

FT TRANSMEM 32 POTENTIAL.

FT TRANSMEM 107 POTENTIAL.

FT TRANSMEM 152 POTENTIAL.

FT TRANSMEM 375 POTENTIAL.

FT TRANSMEM 512 POTENTIAL.

FT TRANSMEM 583 POTENTIAL.

FT TRANSMEM 603 POTENTIAL.

FT TRANSMEM 641 POTENTIAL.

FT TRANSMEM 661 POTENTIAL.

FT TRANSMEM 666 POTENTIAL.

SQ SEQUENCE 865 AA; 99206 MW; 9F7BD97BA4FB1904 CRC64;

Query Match Best Local Similarity 82.5%; Score 33; DB 3; Length 865;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 WQSAFTK 7  
11:1111  
DB 448 WQSAFTK 454.

RESULT 8

O52995 PRELIMINARY: PRT: 77 AA.

AC O52995:

DT 01-JUN-1998 (TREMBlrel. 06, Created)

DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE PCO1D-157 mbd protein.

GN PCO1D-157 MBDD.

```

OS Escherichia coli
OC plasmid pCOLd-157
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CL40 CURED;
RX MEDLINE=98092211; PubMed=9431914;
RA Hofinger C., Karch H., Schmidt H.;
RT "Structure and function of plasmid pCOLd157 of enterohemorrhagic
RT Escherichia coli O157 and its distribution among strains from patients
RT with diarrhea and hemolytic-uremic syndrome.";
RL J. Clin. Microbiol. 36:24-29(1998).
DR EMBL; Y10412; CAA71431.1; -
KW plasmid.
SQ SEQUENCE 77 AA; 9239 MW; 998FB8A10AD8D5CB CRC64;

Query Match 80.0%; Score 32; DB 2; Length 77;
Best Local Similarity 83.3%; Pred. No. 26;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WSAFT 6
   1:||||
Db 26 WSAFT 31

RESULT 9
O92RF9 PRELIMINARY; PRT; 161 AA.
AC O92RF9;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hypothetical peroxiredoxin protein SMC00072.
GN R00915 OR SMC00072
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masny D.,
RA Pohl T., Portetle D., Puehler A., Purnelle B., Ramepenger U.,
RA Renard C., Thebaud P., Vandebol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; AL591785; CAC45487.1; -
DR InterPro: IPR000866; Ahpc-TSA.
DR Pfam; PF00578; Ahpc-TSA; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 161 AA; 16997 MW; 9FAA1CBFD1D31C01 CRC64;

Query Match 80.0%; Score 32; DB 16; Length 161;
Best Local Similarity 71.4%; Pred. No. 57;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 WSAFT 7
   1:||||
Db 102 WSAFT 108

RESULT 10
O98B00 PRELIMINARY; PRT; 369 AA.
AC O98B00;
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)

```

```

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Amino transferase.
GN MFR5786.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL; AP003007; BAB52172.1; -
DR InterPro: IPR004839; Amino transferase/2.
DR InterPro: IPR001917; NMT transferase.
DR Pfam; PF00155; aminotran_1_2; 1.
DR TIGRfams; TIGR01141; hsc; 1.
DR PROSITE; PS00599; AA_TRANSFER_CLASS_2; UNKNOWN_1.
KW transferase; Complete proteome.
SQ SEQUENCE 369 AA; 40710 MW; DCB23E0507165570 CRC64;

Query Match 80.0%; Score 32; DB 16; Length 369;
Best Local Similarity 71.4%; Pred. No. 146+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 WSAFT 7
   1:||||
Db 315 WSAFT 321

RESULT 11
O9UQYO PRELIMINARY; PRT; 385 AA.
AC O9UQYO;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE O-methyltransferase.
GN Mt-I.
OS Aspergillus parasiticus.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic trichocomaceae; Aspergillus.
OX NCBI_TaxID=5067;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NTAH-26;
RX MEDLINE=20011239; PubMed=10543813;
RA Motomura M., Chihaya N., Shinozawa T., Hamasaki T., Yabe K.;
RT "Cloning and characterization of the O-methyltransferase I gene (dmta)
RT from Aspergillus parasiticus associated with the conversions of
RT demethylsterigmatocystin to sterigmatocystin and
RT dihydrodemethylsterigmatocystin to dihydrosterigmatocystin in
RT aflatoxin biosynthesis.";
RL Appl. Environ. Microbiol. 65:4987-4994(1999).
DR EMBL; AB022906; BAA86104.1; -
DR EMBL; AB022905; BAA86103.1; -
DR InterPro: IPR00183; Decarboxylase2.
DR InterPro: IPR001601; Methyltransferase.
DR InterPro: IPR001077; O-methyltransferase.
DR Pfam; PF00891; Methyltransferase_2; 1.
DR PROSITE; PS00879; ODR_DC_2; UNKNOWN_1.
KW Methyltransferase; Transferase.
FT INT MET 0
SQ SEQUENCE 385 AA; 43023 MW; C9866E7A96272005 CRC64;

Query Match 80.0%; Score 32; DB 3; Length 385;

```



Best Local Similarity 83.3%; Pred. No. 1.4e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WOSAF 6  
DB 330 WQAF 335

## RESULT 12

O9UV62 PRELIMINARY; PRT; 386 AA.

AC O9UV62: 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, last sequence update)

DE 01-MAR-2002 (TREMBlrel. 20, last annotation update)

GN OMTB.

OS Aspergillus parasiticus.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

CC Eucotiles; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.

OX NCBI\_Taxid=5067;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-SU-1;

RX MEDLINE-20267858; PubMed=10806361;

RA Yu J., Moloshuk C.P., Bhattacharjee D., Cleveland T.E.;

RT Cloning and characterization of avfa and omtb genes involved in aflatoxin biosynthesis in three Aspergillus species."

RL Gene 248:157-167(2000).

DR EMBL: AF154050; AAF25603.1;

DR InterPro: IPR000183; Decarboxylase2.

DR InterPro: IPR001077; O\_Mettransf2.

DR Pfam: PF00891; Methyltransf\_2; 1.

DR PROSITE: PS00879; ODR\_DC\_2; UNKNOWN\_1.

DR Methyltransferase; Transferase.

DR SEQUENCE 386 AA; 43142 MW; 13D6F1ED63CF63D CRC64;

SO

Query Match

Best Local Similarity 83.3%; Pred. No. 1.4e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WOSAF 6

DB 331 WQAF 336

## RESULT 13

O9P900 PRELIMINARY; PRT; 386 AA.

AC O9P900: 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, last sequence update)

DE 01-MAR-2002 (TREMBlrel. 20, last annotation update)

GN OMTB.

OS Aspergillus flavus.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

CC Eucotiles; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.

OX NCBI\_Taxid=5059;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-CRA01-2B;

RX MEDLINE-20267858; PubMed=10806361;

RA Yu J., Moloshuk C.P., Bhattacharjee D., Cleveland T.E.;

RT Cloning and characterization of avfa and omtb genes involved in aflatoxin biosynthesis in three Aspergillus species."

RL Gene 248:157-167(2000).

DR EMBL: AF159789; AAF26223.1;

DR InterPro: IPR000183; Decarboxylase2.

DR InterPro: IPR001077; O\_Mettransf2.

DR Pfam: PF00891; Methyltransf\_2; 1.

DR PROSITE: PS00879; ODR\_DC\_2; UNKNOWN\_1.

KW Methyltransferase; Transferase.

SO SEQUENCE 386 AA; 43160 MW; F199017CC8D89C1 CRC64;

## Query Match

Best Local Similarity 83.3%; Pred. No. 1.4e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WOSAF 6

DB 331 WQAF 336

## RESULT 14

O9FV0 PRELIMINARY; PRT; 435 AA.

AC O9FV0: 01-MAR-2001 (TREMBlrel. 16, Created)

DT 01-MAR-2001 (TREMBlrel. 16, last sequence update)

DE 01-JUN-2001 (TREMBlrel. 17, last annotation update)

GN Similarity to GTPase activating protein.

OS Arabidopsis thaliana (mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OX eurosids II; Brassicales; Brassicaceae; Arabidopsis.

NCBI\_Taxid=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-COLUMBIA;

RX MEDLINE-97471969; PubMed=9330910;

RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,

RT Miyajima N., Tabata S.;

RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence features of the 1.6 Mb regions covered by twenty physically assigned

RT P1 clones."

RL DNA Res. 4:215-230(1997).

DR EMBL: AB005232; BAB08757.1;

DR InterPro: IPR00195; RabGAP\_TBC.

DR Pfam: PF00566; TBC; 1.

DR SMART: SM00164; TBC; 1.

DR SEQUENCE 435 AA; 49768 MW; A967D4DC4F703C07 CRC64;

## Query Match

Best Local Similarity 83.3%; Pred. No. 1.6e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WOSAF 6

DB 40 WQAF 45

## RESULT 15

O9GRW8 PRELIMINARY; PRT; 473 AA.

AC O9GRW8: 01-MAR-2001 (TREMBlrel. 16, Created)

DT 01-MAR-2001 (TREMBlrel. 16, last sequence update)

DE 01-JUN-2001 (TREMBlrel. 17, last annotation update)

GN GpII protein.

OS Drosophila virilis (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_Taxid=7244;

RN [1]

RP SEQUENCE FROM N.A.

RA Campese S., Chalmers D., Sandrelli F., Megighian A., Pelxoto A.A.,

RT Costa R., Kyriacou C.P.;

RT "Comparative analysis of the nonA region in Drosophila identifies

RT highly diverged 5' gene that may constrain nonA promoter evolution."

RL Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.

DR EMBL: AJ298996; CAC10057.1;

DR FlyBase; FBgn0042731; Dvir\Gp11.  
SQ SEQUENCE 473 AA; 55085 MW; A6E3E375C5AB348A CRC64;

Query Match 80.0%; Score 32; DB 5; Length 473;  
Best Local Similarity 71.4%; Pred. No. 1.8e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 WQSAFTR 7  
111 : 11  
Db 202 WQSIYTR 208

Search completed: July 11, 2003, 11:58:51  
Job time : 6.05613 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:42:16 ; Search time 10.9023 seconds  
(without alignments)  
232.223 Million cell updates/sec

Title: US-10-091-442-3

Perfect score: 100

Sequence: 1 VERVDFTHLEDTRRNINK 19

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A.GeneSeq\_101002:\*

1: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT:\*

2: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT:\*

3: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT:\*

4: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT:\*

5: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT:\*

6: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT:\*

7: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT:\*

8: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT:\*

9: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT:\*

10: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT:\*

11: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT:\*

12: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT:\*

13: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT:\*

14: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT:\*

15: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT:\*

16: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT:\*

17: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT:\*

18: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT:\*

19: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT:\*

20: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT:\*

21: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT:\*

22: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT:\*

23: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	100	100.0	380	15	AA848379 Human megakaryocyte
2	100	100.0	380	20	AAV08254 Human megaslin prote
3	100	100.0	380	21	AA824142 Human megaslin prote
4	100	100.0	380	22	AA83075 Human megaslin prote
5	82	82.0	19	15	AA857109 Human megakaryocyte
6	71	71.0	368	20	AAV08256 Mouse megaslin prote
7	71	71.0	368	21	AA824151 Mouse megaslin prote
8	71	71.0	368	22	AA83077 Murine megaslin prot
9	68	68.0	380	20	AAV08255 Rat megaslin protein
10	68	68.0	380	21	AA824150 Rat megaslin protein

11	68	68.0	380	22	AA64286 Rat megaslin protein
12	68	68.0	380	22	AA83076 Rat megaslin protein
13	50	50.0	390	13	AA825276 SCC antigen. Synt
14	50	50.0	390	18	AA815241 Psoriasis-like type
15	50	50.0	390	20	AAV32077 Hepatitis B virus
16	50	50.0	390	20	AAV32078 Hepatitis B virus
17	50	50.0	390	20	AAV25928 Human SCCA1 protei
18	50	50.0	390	22	AAV72654 Human squamous cel
19	47	47.0	482	16	AA872601 Truncated human ch
20	47	47.0	503	16	AA872508 Rat cholesterol 7
21	47	47.0	504	13	AA82482 Cholesterol 7-alpha
22	47	47.0	504	16	AA872507 Human cholesterol
23	47	47.0	504	16	AA872509 Hamster cholesterol
24	47	47.0	504	16	AA872505 Human cholesterol
25	47	47.0	1047	22	AA872507 Human protein SRO
26	47	47.0	1047	22	AA872507 Human protein SRO
27	46	46.0	375	15	AA850938 Mespin. Homo sapi
28	46	46.0	375	15	AA850938 Tumour suppressor
29	46	46.0	375	19	AA869287 Mespin. Homo sapi
30	46	46.0	375	20	AAV02607 Human mespin prote
31	46	46.0	390	18	AA815242 Psoriasis-like type
32	46	46.0	390	20	AAV25927 Human SCCA2 protei
33	46	46.0	390	22	AAV72655 Human squamous cel
34	46	46.0	410	22	AA804102 Botulinum toxin hea
35	46	46.0	427	22	AB867060 Drosophila melanog
36	46	46.0	579	23	AB897393 Novel human protei
37	46	46.0	732	17	AA800466 Bovine Factor-XIII
38	46	46.0	732	18	AA836105 Wild type bovine F
39	46	46.0	732	18	AA836106 Mutant bovine fact
40	46	46.0	1120	22	ABG28279 Novel human diagno
41	45	45.0	45	22	AB843663 Peptide #1169 enco
42	45	45.0	45	22	AB826613 Protein #8612 enco
43	45	45.0	45	22	AA864615 Human brain expres
44	45	45.0	45	22	AA877400 Human bone marrow
45	45	45.0	45	22	AA821329 Peptide #763 enco

#### ALIGNMENTS

RESULT 1	AA848379	standard; Protein; 380 AA.
ID	AA848379	
AC	AA848379	
DT	16-AUG-1994	(first entry)
DE	Human megakaryocyte differentiation factor.	
XX	Human megakaryocyte differentiation factor; MDF; thrombopoietin;	
KW	haematopoietic stimulating factor; thrombocytopoiesis; platelet;	
KW	bone marrow transplantation; cancer chemotherapy.	
OS	Homo sapiens.	
PN	EP583884-A.	
PD	23-FEB-1994.	
PF	19-JUL-1993; 93BP-0305654.	
PR	17-JUL-1992; 92JP-0212305.	
PR	04-MAR-1993; 93JP-0067339.	
PA	(SUNR) SUNTORY LTD.	
PA	(TSUJ) TSUJIMOTO M.	
PI	Ishida N, Iwasa F, Kurihara T, Miura K, Nakazato H;	
PI	Tsujimoto M, Tsunokuchi N, Yamaguchi N, Yamachi K;	
DR	WPI: 1994-058782/08.	
DR	N-PSDB: AA056670.	

```

XX  New megakaryocyte differentiation factor - isolated from human
PT  epidermoid carcinoma cells, used to treat conditions involving a
PT  decrease in platelets
XX
PS  Claim 7; Page 30-32; 47pp; English.
XX
CC  Human MDF can be isolated from a culture of human epidermoid
CC  carcinoma A431 cells in protein-free medium. The MDF stimulates
CC  differentiation of megakaryocytes from myeloid cells in the presence
CC  of IL-3. The MDF acts in vivo as a thrombopoietin making it useful
CC  for treatment of diseases involving a decrease in platelet number
CC  (esp. thrombocytopenia) such as occurs in bone marrow
CC  transplantation and in chemotherapy. MDF has mol.wt. 55-57kd by
CC  SDS-PAGE and exhibits an isoelectric point of 6.0-7.0.
XX
SQ  Sequence 380 AA;

Query Match 100.0%; Score 100; DB 15; Length 380;
Best Local Similarity 100.0%; Pred. No. 7.6e-08;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY  1 VERVDFTNHLEDTRRNINK 19
DB  126 VERVDFTNHLEDTRRNINK 144

RESULT 2
AAy08254
ID  AAy08254 standard; Protein: 380 AA.
XX
AC  AAY08254;
XX
DT  14-JUL-1999 (first entry)
XX
DE  Human megsin protein.
XX
KW  Megsin; mesangial cell; treatment; diagnosis; disease; IGA nephropathy;
KW  human; rat; murine.
XX
OS  Homo sapiens.
XX
PN  W09915652-A1.
XX
PD  01-APR-1999.
XX
PF  22-SEP-1998; 98WO-JP04269.
XX
PR  22-SEP-1997; 97JP-0275302.
XX
PA  (KURO/) KUROKAWA K.
PA  (MIYA/) MIYATA T.
XX
PI  Miyata T;
XX
DR  WPI; 1999-276983/23.
DR  N-PSDB; AAX56712.
XX
PT  Megsin protein expressed specifically in mesangial cells
XX
PS  Claim 1; Page 62-64; 100pp; Japanese.
XX
CC  This invention describes the isolation of novel megsin nucleic acid and
CC  proteins from human, rat and mouse tissue. This protein is expressed
CC  specifically in mesangial cells. The products of the invention are
CC  useful for the treatment and diagnosis of diseases involving mesangial
CC  cells, such as IGA nephropathy.
XX
SQ  Sequence 380 AA;

Query Match 100.0%; Score 100; DB 20; Length 380;
Best Local Similarity 100.0%; Pred. No. 7.6e-08;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY  1 VERVDFTNHLEDTRRNINK 19
DB  126 VERVDFTNHLEDTRRNINK 144

RESULT 3
AAB24142
ID  AAB24142 standard; Protein: 380 AA.
XX
AC  AAB24142;
XX
DT  30-JAN-2001 (first entry)
XX
DE  Human megsin protein sequence SEQ ID NO:2.
XX
KW  Megsin; mesangium-predominant gene; seipin regulated; nephropathy;
KW  IGA; immunoglobulin A; detection; renal function; renal disorder;
KW  diagnosis; biological sample; blood; urine.
XX
OS  Homo sapiens.
XX
PN  W0200057189-A1.
XX
PD  28-SEP-2000.
XX
PF  17-MAR-2000; 2000WO-JP01646.
XX
PR  19-MAR-1999; 99JP-0075305.
PR  28-OCT-1999; 99JP-0306623.
XX
PA  (KURO/) KUROKAWA K.
PA  (FUSO ) FUSO PHARM IND LTD.
PA  (MIYA/) MIYATA T.
XX
PI  Miyata T;
XX
DR  WPI; 2000-611642/58.
DR  N-PSDB; AAB9294.
XX
PT  Evaluating renal function comprises assaying megsin protein in
PT  biological sample
XX
PS  Example 2; Page 66-69; 93pp; Japanese.
XX
CC  The present invention describes a method for evaluating renal function.
CC  The method comprises assaying megsin protein in biological sample. Also
CC  described are: (1) use of a anti-megsin protein antibody for diagnosing
CC  renal function; and (2) a kit for detecting megsin protein comprising:
CC  (a) anti-megsin protein antibody attached to solid magnetic particles;
CC  (b) direct or indirect fixing for the antibody to the particles; and
CC  (c) a magnet. The process is useful for evaluating renal function and
CC  diagnosing renal disorders by assaying megsin protein in biological
CC  samples (preferably urine or blood). The process is reproducible and
CC  gives accurate results. The present sequence represents the human megsin
CC  protein, which is given in the exemplification of the present invention.
XX
SQ  Sequence 380 AA;

Query Match 100.0%; Score 100; DB 21; Length 380;
Best Local Similarity 100.0%; Pred. No. 7.6e-08;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY  1 VERVDFTNHLEDTRRNINK 19
DB  126 VERVDFTNHLEDTRRNINK 144

RESULT 4
AAB83075
ID  AAB83075 standard; Protein: 380 AA.
XX
AC  AAB83075;

```

```

XX 10-JUL-2001 (first entry)
DT Human megasin protein.
XX
DE Human; megasin; mesangial cell proliferative nephritis; nephrotropic;
XX transgenic mouse; glomerular disease; animal model; drug screening.
XX
OS Homo sapiens.
XX
PN WO200124628-A1.
XX
PD 12-APR-2001.
XX
PF 06-OCT-2000; 2000WO-JP06988.
XX
PR 06-OCT-1999; 99JP-0285736.
XX
PA (KURO/) KUROKAWA K.
XX (MIYA/) MIYATA T.
XX
PI Miyata T;
XX
DR WPI: 2001-300136/31.
XX N-PSDB: AAF82438.
XX
PT Mouse model for mesangial cell proliferative nephritis for development
XX and screening of new treatments.
XX
PS Example 4; Page 44-46; 62pp; Japanese.
XX
CC The present sequence is human megasin. The human megasin coding
XX sequence may be introduced into a mouse to produce an animal model of
XX mesangial cell proliferative nephritis. The symptoms include
XX enlargement of the mesangial base region, sedimentation of an immune
XX complex and an increase in mesangial cells. The animal model is useful
XX for analysing the pathology of chronic glomerular diseases and for
XX screening compositions for prevention and treatment of the diseases.
XX Highly uniform models can be made easily and in large numbers using
XX this method.
XX
SQ Sequence 380 AA:
XX
Query Match 100.0%; Score 100; DB 22; Length 380;
Best Local Similarity 100.0%; Pred. No. 7,6e-08;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 VERVDFTNHLNEDTRRNINK 19
DB 126 VERVDFTNHLNEDTRRNINK 144

```

```

FT Misc-difference 15 /note= "not determined"
FT
XX EP583884-A.
XX
PN 23-FEB-1994.
XX
PD 19-JUL-1993; 93EP-0305654.
XX
PF 17-JUL-1992; 92JP-0212305.
XX PR 04-MAR-1993; 93JP-0067339.
XX
XX (SUNR ) SUNTORY LTD.
XX (TSUJ/) TSUJIMOTO M.
XX
XX Ishida N, Iwasa F, Kurihara T, Miura K, Nakazato H;
XX Tsujimoto M, Tsuruoka N, Yamaguchi N, Yamachi K;
XX WPI: 1994-058782/08.
XX
DR
XX
PT New megakaryocyte differentiation factor - isolated from human
XX epidermoid carcinoma cells, used to treat conditions involving a
XX decrease in platelets
XX
PS Claim 1; Page 18; 47pp; English.
XX
CC Human MDF (see AAR48379) can be isolated from a culture of human
XX epidermoid carcinoma A431 cells in protein-free medium. The MDF
XX stimulates differentiation of megakaryocytes from myeloid cells
XX in the presence of IL-3. The MDF acts in vivo as a thrombopoietin
XX making it useful for treatment of diseases involving a decrease
XX in platelet number (esp. thrombocytopenia) such as occurs in bone
XX marrow transplantation and in chemotherapy. MDF has mol.wt.
XX 55-57kD by SDS-PAGE, exhibits an isoelectric point of 6.0-7.0 and
XX contains an amino acid sequence comprising at least one of the
XX sequences AAR57107-R57115.
XX
SQ Sequence 19 AA:
XX
Query Match 82.0%; Score 82; DB 15; Length 19;
Best Local Similarity 84.2%; Pred. No. 1.8e-06;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 VERVDFTNHLNEDTRRNINK 19
DB 1 VEXVDFTNHLNEDTXNINK 19

```

```

RESULT 5
AAR57109
ID AAR57109 standard; peptide; 19 AA.
XX
AC AAR57109;
XX
DE 16-AUG-1994 (first entry)
XX
XX Human megakaryocyte differentiation factor peptide 3.
XX
XX Human megakaryocyte differentiation factor; MDF; thrombopoietin;
XX haematopoietic stimulating factor; thrombocytopenia; platelet;
XX bone marrow transplantation; cancer chemotherapy.
XX
OS Homo sapiens.
XX
XX
XX Key Location/Qualifiers
XX FT Misc-difference 3 /note= "not determined"
XX FT Misc-difference 14 /note= "not determined"
XX FT

```

```

RESULT 6
AAV08256
ID AAV08256 standard; Protein; 368 AA.
XX
AC AAV08256;
XX
DE 14-JUL-1999 (first entry)
XX
XX Mouse megasin protein.
XX
XX Megasin; mesangial cell; treatment; diagnosis; disease; IgA nephropathy;
XX human; rat; murine.
XX
XX Mus musculus.
XX
XX Key Location/Qualifiers
XX FT Protein 1..368
XX FT /note= "partial sequence"
XX
XX WO9915652-A1.
XX
XX 01-APR-1999.
XX
XX 22-SEP-1998; 98WO-JP04269.
XX

```

PR 22-SEP-1997; 97JP-0275302.  
XX  
XX (KURO/) KUROKAWA K.  
PA (MIYA/) MIYATA T.  
XX  
XX  
PI Miyata T;  
XX  
XX WPI: 1999-276983/23.  
DR N-PSDB; AAX56714.  
XX  
XX  
PT Megin protein expressed specifically in mesangial cells  
PS  
PS Claim 1; Page 76-79; 100pp; Japanese.  
XX  
XX This invention describes the isolation of novel megin nucleic acid and  
CC proteins from human, rat and mouse tissue. This protein is expressed  
CC specifically in mesangial cells. The products of the invention are  
CC useful for the treatment and diagnosis of diseases involving mesangial  
CC cells, such as Iga nephropathy.  
XX  
XX  
SQ Sequence 368 AA;  
  
Query Match 71.0%; Score 71; DB 20; Length 368;  
Best Local Similarity 73.7%; Pred. No. 0.0034;  
Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
  
OY 1 VERVDFTNHLEDTRRNINK 19  
DB 114 VERVDFTNVDVDFTRFKINK 132  
|||||||:|||||  
  
RESULT 7  
AAB24151  
ID AAB24151 standard; Protein: 368 AA.  
XX  
AC AAB24151;  
XX  
DT 30-JAN-2001 (first entry)  
XX  
DE Mouse megin protein sequence SEQ ID NO:21.  
XX  
XX Megin; mesangium-predominant gene; serpin regulated; nephropathy;  
KW Iga; immunoglobulin A; detection; renal function; renal disorder;  
KM diagnosis; biological sample; blood; urine.  
XX  
OS Mus musculus.  
XX  
XX WO200057189-A1.  
PN  
XX 28-SEP-2000.  
PD  
XX 17-MAR-2000; 2000WO-JP01646.  
PF  
XX 19-MAR-1999; 99JP-0075305.  
PR 28-OCT-1999; 99JP-0306623.  
XX  
XX (KURO/) KUROKAWA K.  
PA (FUSO ) FUSO PHARM IND LTD.  
PA (MIYA/) MIYATA T.  
XX  
XX Miyata T;  
PI  
XX WPI: 2000-611642/58.  
DR N-PSDB; AAC55239.  
XX  
XX  
PT Evaluating renal function comprises assaying megin protein in  
PT biological sample  
XX  
XX Disclosure; Page 89-91; 93pp; Japanese.  
PS  
XX The present invention describes a method for evaluating renal function.  
CC The method comprises assaying megin protein in biological sample. Also  
CC described are: (1) use of a anti-megin protein antibody for diagnosing

CC renal function; and (2) a kit for detecting megin protein comprising:  
CC (a) anti-megin protein antibody attached to solid magnetic particles;  
CC (b) direct or indirect fixing for the antibody to the particles; and  
CC (c) a magnet. The process is useful for evaluating renal function and  
CC diagnosing renal disorders by assaying megin protein in biological  
CC samples (preferably urine or blood). The process is reproducible and  
CC gives accurate results. The present sequence represents the mouse megin  
CC protein, which is given in the exemplification of the present invention.  
XX  
XX  
SQ Sequence 368 AA;  
  
Query Match 71.0%; Score 71; DB 21; Length 368;  
Best Local Similarity 73.7%; Pred. No. 0.0034;  
Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
  
OY 1 VERVDFTNHLEDTRRNINK 19  
DB 114 VERVDFTNVDVDFTRFKINK 132  
|||||||:|||||  
  
RESULT 8  
AAB83077  
ID AAB83077 standard; Protein: 368 AA.  
XX  
AC AAB83077;  
XX  
DT 10-JUL-2001 (first entry)  
XX  
DE Murine megin protein.  
XX  
XX Mouse; megin; mesangial cell proliferative nephritis; nephrotropic;  
KW transgenic mouse; glomerular disease; animal model; drug screening.  
XX  
XX Mus musculus.  
XX  
XX WO200124628-A1.  
PN  
XX 12-APR-2001.  
PD  
XX 06-OCT-2000; 2000WO-JP06988.  
PF  
XX 06-OCT-1999; 99JP-0285736.  
PR  
XX (KURO/) KUROKAWA K.  
PA (MIYA/) MIYATA T.  
XX  
XX Miyata T;  
PI  
XX WPI: 2001-300136/31.  
DR N-PSDB; AAF82440.  
XX  
XX Mouse model for mesangial cell proliferative nephritis for development  
PT and screening of new treatments  
XX  
XX Disclosure; Page 52-53; 62pp; Japanese.  
PS  
XX The present sequence is murine megin protein. The human megin coding  
CC sequence may be introduced into a mouse to produce an animal model of  
CC mesangial cell proliferative nephritis. The symptoms include  
CC enlargement of the mesangial base region, sedimentation of an immune  
CC complex and an increase in mesangial cells. The animal model is useful  
CC for analysing the pathology of chronic glomerular diseases and for  
CC screening compositions for prevention and treatment of the diseases.  
CC Highly uniform models can be made easily and in large numbers using  
CC this method.  
XX  
XX  
SQ Sequence 368 AA;  
  
Query Match 71.0%; Score 71; DB 22; Length 368;  
Best Local Similarity 73.7%; Pred. No. 0.0034;  
Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
  
OY 1 VERVDFTNHLEDTRRNINK 19

```

Db      114 VERVDFTNDVQDTRFKINK 132
      |||||||::||| |||
RESULT 9
ID      AAY08255
AC      AAY08255;
XX      14-JUL-1999 (first entry)
DT      Rat megalin protein.
DE      Megalin: mesangial cell; treatment; diagnosis; disease; Iga nephropathy;
KW      human; rat; murine.
XX      Rattus rattus.
OS      WO915652-A1.
XX      01-APR-1999.
PD      22-SEP-1998; 98WO-JP04269.
XX      22-SEP-1997; 97JP-0275302.
PR      (KUROO/) KUROKAWA K.
XX      (MIYA/) MIYATA T.
PA      Miyata T;
XX      WPI: 1999-276983/23.
DR      N-PSDB; AAX56712.
XX      Megalin protein expressed specifically in mesangial cells
PT      Claim 1; Page 69-72; 100pp; Japanese.
XX      This invention describes the isolation of novel megalin nucleic acid and
CC      proteins from human, rat and mouse tissue. This protein is expressed
CC      specifically in mesangial cells. The products of the invention are
CC      useful for the treatment and diagnosis of diseases involving mesangial
CC      cells, such as Iga nephropathy.
XX      Sequence 380 AA:
SQ
Query Match      68.0%; Score 68; DB 20; Length 380;
Best Local Similarity 68.4%; Pred. No. 0.011;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
OY      1 VERVDFTNHLDETRRINK 19
      |||||||::||| |||
Db      126 VERVDFTNDIOETRFKINK 144
      |||||||::||| |||
RESULT 10
ID      AAB24150
AC      AAB24150;
XX      30-JAN-2001 (first entry)
DT      Rat megalin protein sequence SEQ ID NO:19.
DE      Megalin: mesangial-predominant gene; serpin regulated; nephropathy;
KW      Iga; immunoglobulin A; detection; renal function; renal disorder;
XX      diagnosis; biological sample; blood; urine.
OS      Rattus norvegicus.
XX      Key      Location/Qualifiers
FH

```

```

FT      Misc-difference 51
FT      /note= "unspecified"
FT      Misc-difference 94
FT      /note= "unspecified"
XX      WO200057189-A1.
XX      28-SEP-2000.
XX      17-MAR-2000; 2000WO-JP01646.
XX      19-MAR-1999; 99JP-0075305.
PR      28-OCT-1999; 99JP-0306623.
XX      (KUROO/) KUROKAWA K.
PA      (FUSO) FUSO PHARM IND LTD.
XX      (MIYA/) MIYATA T.
PI      Miyata T;
XX      WPI: 2000-611642/58.
DR      N-PSDB; AAC35238.
XX      Evaluating renal function comprises assaying megalin protein in
PT      biological sample -
XX      Example 2; Page 81-84; 93pp; Japanese.
XX      The present invention describes a method for evaluating renal function.
CC      The method comprises assaying megalin protein in biological sample. Also
CC      described are: (1) use of an anti-megalin protein antibody for diagnosing
CC      renal function; and (2) a kit for detecting megalin protein comprising:
CC      (a) anti-megalin protein antibody attached to solid magnetic particles;
CC      (b) direct or indirect fixing for the antibody to the particles; and
CC      (c) a magnet. The process is useful for evaluating renal function and
CC      diagnosing renal disorders by assaying megalin protein in biological
CC      samples (preferably urine or blood). The process is reproducible and
CC      gives accurate results. The present sequence represents the rat megalin
CC      protein, which is given in the exemplification of the present invention.
XX      Sequence 380 AA:
SQ
Query Match      68.0%; Score 68; DB 21; Length 380;
Best Local Similarity 68.4%; Pred. No. 0.011;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
OY      1 VERVDFTNHLDETRRINK 19
      |||||||::||| |||
Db      126 VERVDFTNDIOETRFKINK 144
      |||||||::||| |||
RESULT 11
ID      AAG64286
AC      AAG64286;
XX      21-SEP-2001 (first entry)
DT      Rat megalin protein.
DE      Rat; megalin; renal mesangial cell; mesangium proliferative nephritis.
XX      Rattus norvegicus.
OS      WO200148019-A1.
XX      05-JUL-2001.
XX      26-DEC-2000; 2000WO-JP09251.
PF      28-DEC-1999; 99JP-0373677.
XX

```





PT New DNA encoding squamous cell carcinoma antigen - allowing large  
 PT scale antigen prodn. for use as diagnostic reagent or immunogen,  
 PT also useful as hybridisation probe  
 PS Claim 2: Page 9-10, Fig 4; 21pp; German.  
 XX  
 CC The protein relates to an antigen (Ag) associated with squamous cell  
 CC carcinoma (SCC). The Ag (which can be isolated from SCC of the  
 CC cervix uteri, pulmonary or oesophageal carcinoma, etc.) is known as a  
 CC diagnostic reagent for SCC but is not easily produced on a large scale  
 CC from human cancer cells. By expressing the DNA sequence encoding this  
 CC protein, in e.g. E.coli, yeast or mammalian cells, Ag can be prep'd. In  
 CC large quantities. Ag can also be used as an immunogen (to produce  
 CC monoclonal antibodies, also useful diagnostically) while the Ag-  
 CC encoding gene can be detected by hybridisation with this sequence.  
 CC Ag is homologous with serine protease inhibitors so may be useful  
 CC therapeutically.  
 SQ Sequence 390 AA;  
 Query Match 50.0%; Score 50; DB 13; Length 390;  
 Best Local Similarity 55.6%; Pred. No. 8.6;  
 Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
 OY 1 VERVDFTNHLDETRRNIN 18  
 DB 131 VESVDFANAPESRRKIN 148  
 RESULT 14  
 AAW15241  
 ID AAW15241 standard; Protein; 390 AA.  
 AC AAW15241;  
 DT 21-JUL-1997 (first entry)  
 DE Psoriasis type I.  
 DE Psoriasis type I.  
 KW Psoriasis; psoriasis; cancer; cell proliferation; therapy;  
 KW antagonist.  
 OS Homo sapiens.  
 OS  
 FH Key Location/Qualifiers  
 FT Active-site 357  
 FT /note= "Thr-357 may correspond to the P'2 or P'3  
 FT position of reactive site of homologous  
 FT serpins"  
 FT  
 PN MO9714425-A1.  
 PD 24-APR-1997.  
 PF 16-OCT-1996; 96WO-US16599.  
 PR 20-MAR-1996; 96US-0013755.  
 PR 17-OCT-1995; 95US-0005679.  
 PA (GEHO ) GEN HOSPITAL CORP.  
 PA (SHIS ) SHISEIDO CO LTD.  
 PI Bacu PC, Goetlinck PF, Hibino T, Takahashi T;  
 DR N-PSDB: AAT65948.  
 DR WPI: 1997-244853/22.  
 PT Modulating cell proliferation or apoptosis - by modulating activity  
 PT of psoriasis type I and II genes, e.g. using (ant)agonists,  
 PT useful for treatment of cancer or psoriasis  
 PS Disclosure: Page 44-45; 63pp; English.  
 CC Psoriasis type I (AAW15241) is a 43 kDa polypeptide that is

CC expressed in psoriatic tissue, e.g. psoriatic epidermis, at least 2  
 CC and pref. at least 5-10 times more abundantly than in normal  
 CC tissue. It is a cross-class inhibitor, inhibits cathepsin L less  
 CC efficiently than does squamous cell carcinoma-antigen (SCC-A), does  
 CC not inhibit cathepsin B or cathepsin H, is active at pH 5.0, and is  
 CC secreted. Its amino acid sequence, deduced from an isolated cDNA  
 CC clone (AAT65948), shows 98% homology to SCC-A. A related sequence,  
 CC psoriasis type II (AAW15242), has also been identified. These  
 CC polypeptides can be used to identify psoriasis antagonists useful  
 CC in the treatment of cell proliferation disorders, e.g. cancer and  
 CC psoriasis.  
 SQ Sequence 390 AA;  
 Query Match 50.0%; Score 50; DB 18; Length 390;  
 Best Local Similarity 55.6%; Pred. No. 8.6;  
 Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
 OY 1 VERVDFTNHLDETRRNIN 18  
 DB 131 VESVDFANAPESRRKIN 148  
 RESULT 15  
 AAY32077  
 ID AAY32077 standard; Protein; 390 AA.  
 AC AAY32077;  
 DT 17-JAN-2000 (first entry)  
 DE Hepatitis B virus receptor SCCL.  
 DE Hepatitis B virus receptor SCCL.  
 KW SCCL; squamous cell carcinoma antigen; human; hepatitis B virus;  
 KW HBV; receptor; antiviral; virucide; screening; therapy.  
 OS Homo sapiens.  
 OS  
 PN EP952217-A2.  
 PD 27-OCT-1999.  
 PF 03-MAR-1999; 99EP-0200603.  
 PR 12-MAR-1998; 98IT-MI00498.  
 PA (ITUY-) ITAL MIN UNIV RICERCA SCI & TECNOLOGICA.  
 PI Fassina G, De Falco S, Verdoliva A, Ruvo M;  
 DR WPI: 1999-602902/52.  
 DR N-PSDB: AAZ20257.  
 PT Use of a recombinant protein and its allelic variant as receptors for a  
 PT hepatitis virus, useful for producing and screening for antiviral drugs  
 CC  
 CC Claim 1: Fig 1; 13pp; English.  
 CC The present sequence comprises human squamous cell carcinoma  
 CC antigen (SCCA1), a protein newly identified as being a receptor for  
 CC hepatitis B virus (HBV). Recombinant proteins having fully, or  
 CC in part, the present sequence or an allelic variant (see AAY32078)  
 CC of it, are claimed and can be produced in claimed prokaryotic  
 CC or eukaryotic (especially human HepG2 cell line) host cells  
 CC using claimed vectors carrying SCCA1 nucleic acids (see AAZ20257 and  
 CC AAZ20258). Transgenic animals expressing SCCA1 or its allelic  
 CC variants are also claimed. The recombinant SCCA1 proteins are used  
 CC in molecular screening assays to identify new drugs, as drugs for  
 CC the treatment of hepatitis B and in the production of monoclonal  
 CC and polyclonal antibodies (claimed). The transgenic animals allow  
 CC the possibility of in vivo evaluation of the effect of compounds  
 CC with anti-hepatitis activity.

XX Sequence 390 AA;

Query Match 50.0%; Score 50; DB 20; Length 390;  
 Best Local Similarity 55.6%; Pred. No. 8.6;  
 Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 VERVDFTNHLEDTRRNIN 18  
 || || | : : | : ||  
 Db 131 VESYDFANAPESRKRIN 148

Search completed: July 11, 2003, 11:54:19  
 Job time : 11.9023 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:49:36 ; Search time 3.39709 Seconds  
(Without alignments)  
164.563 Million cell updates/sec

Title: US-10-091-442-3

Perfect score: 100

Sequence: 1 VERVEFTNHLIEDTRRNINK 19

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_patents\_AA: \*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep: \*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep: \*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep: \*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep: \*  
5: /cgn2\_6/ptodata/1/1aa/PTCUS.COMB.pep: \*  
6: /cgn2\_6/ptodata/1/1aa/backfile1.pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	100	100.0	19	2	US-08-611-977-3
2	100	100.0	380	2	US-08-472-659-34
3	100	100.0	380	2	US-08-474-661-34
4	100	100.0	380	2	US-08-611-977-34
5	82	82.0	19	2	US-08-472-659-3
6	82	82.0	19	2	US-08-474-661-3
7	50	50.0	390	1	US-08-568-147B-2
8	47	47.0	482	1	US-08-135-510-5
9	47	47.0	482	1	US-08-483-852-5
10	47	47.0	482	1	US-08-477-952-5
11	47	47.0	482	2	US-08-477-952-5
12	47	47.0	503	1	US-08-135-511-29
13	47	47.0	503	1	US-08-187-453-29
14	47	47.0	504	1	US-08-135-511-28
15	47	47.0	504	1	US-08-135-511-30
16	47	47.0	504	1	US-08-483-852-8
17	47	47.0	504	1	US-08-361-458-3
18	47	47.0	504	1	US-08-477-953-8
19	47	47.0	504	1	US-08-187-453-28
20	47	47.0	504	1	US-08-187-453-30
21	47	47.0	504	2	US-08-477-952-8
22	47	47.0	506	2	US-08-845-161A-18
23	47	47.0	506	4	US-09-270-751-18
24	46	46.0	375	1	US-08-121-714-2
25	46	46.0	375	1	US-08-477-108A-2
26	46	46.0	375	2	US-08-477-112-2
27	46	46.0	375	5	PCT-US93-08322-2

28	46	46.0	732	5	PCT-US95-17026-2
29	45	45.0	391	4	US-09-123-912-110
30	45	45.0	391	4	US-09-643-597-110
31	45	45.0	400	4	US-09-123-912-112
32	45	45.0	400	4	US-09-643-597-112
33	43	43.0	390	4	US-09-266-910-3
34	43	43.0	390	4	US-09-266-910-4
35	43	43.0	443	4	US-09-134-001C-4915
36	42	42.0	261	4	US-09-110-959A-2
37	42	42.0	317	4	US-09-134-001C-4537
38	41	41.0	376	4	US-09-200-965-2
39	41	41.0	379	1	US-08-121-714-4
40	41	41.0	379	1	US-08-477-108A-4
41	41	41.0	379	2	US-08-477-112-4
42	41	41.0	379	5	PCT-US93-08322-4
43	41	41.0	501	4	US-09-134-001C-5280
44	40.5	40.5	616	4	US-08-965-902A-2
45	40	40.0	279	3	US-08-872-979-5

#### ALIGNMENTS

RESULT 1  
US-08-611-977-3  
Sequence 3, Application US/06611977  
Patent No. 5972886  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSUROOKA, No. 5972886uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. 5972886uhro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAUCHI, Kozo  
APPLICANT: YAMAGUCHI, No. 5972886oml  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/611,977  
FILING DATE: 06-MAR-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-204  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:

Sequence 2, Appl1  
Sequence 110, App  
Sequence 110, App  
Sequence 112, App  
Sequence 3, Appl1  
Sequence 4, Appl1  
Sequence 4915, Ap  
Sequence 2, Appl1  
Sequence 4537, Ap  
Sequence 2, Appl1  
Sequence 4, Appl1  
Sequence 4, Appl1  
Sequence 4, Appl1  
Sequence 5280, Ap  
Sequence 2, Appl1  
Sequence 5, Appl1

LENGTH: 19 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-611-977-3

Query Match 100.0%; Score 100; DB 2; Length 19;  
Best Local Similarity 100.0%; Pred. No. 1.6e-10;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VERVDFTNHLEDTRRNINK 19  
Db 1 VERVDFTNHLEDTRRNINK 19

## RESULT 2

US-08-472-659-34  
Sequence 34, Application US/08472659  
Patent No. 5831030  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IMASA, Fuyuki  
APPLICANT: TSURUOKA, No. 5831030uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. 5831030uhiro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, No. 5831030omi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/472,659  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-248  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-472-659-34

Query Match 100.0%; Score 100; DB 2; Length 380;

Best Local Similarity 100.0%; Pred. No. 5.7e-09;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VERVDFTNHLEDTRRNINK 19  
Db 126 VERVDFTNHLEDTRRNINK 144

## RESULT 3

US-08-474-661-34  
Sequence 34, Application US/08474661  
Patent No. 5874253  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IMASA, Fuyuki  
APPLICANT: TSURUOKA, No. 5874253uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. 5874253uhiro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, No. 5874253omi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/474,661  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: REA, TERESA STANER  
REGISTRATION NUMBER: 30,427  
REFERENCE/DOCKET NUMBER: 001560-204  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-6620  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-474-661-34

Query Match 100.0%; Score 100; DB 2; Length 380;  
Best Local Similarity 100.0%; Pred. No. 5.7e-09;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VERVDFTNHLEDTRRNINK 19  
Db 126 VERVDFTNHLEDTRRNINK 144

## RESULT 4

US-08-611-977-34  
; Sequence 34, Application US/08611977  
; Patent No. 5972886  
; GENERAL INFORMATION:  
; APPLICANT: TSUJIMOTO, Masafumi  
; APPLICANT: IWASA, Fuyuki  
; APPLICANT: TSUROOKA, No. 5972886uo  
; APPLICANT: NAKAZATO, Hiroshi  
; APPLICANT: MIURA, Kenju  
; APPLICANT: ISHIDA, No. 5972886uhltro  
; APPLICANT: KURIHARA, Tatsuya  
; APPLICANT: YAMAGUCHI, Kozo  
; APPLICANT: YAMAGUCHI, No. 5972886oml  
; TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
; NUMBER OF SEQUENCES: 34  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: P.O. Box 1404  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/611.977  
; FILING DATE: 06-MAR-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/091,028  
; FILING DATE: 14-JUL-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 4-212305  
; FILING DATE: 17-JUL-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 6-067339  
; FILING DATE: 04-MAR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McGowan, Malcolm K.  
; REGISTRATION NUMBER: 39,300  
; REFERENCE/DOCKET NUMBER: 001560-204  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 34:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 380 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-611-977-34  
Query Match 100.0%; Score 100; DB 2; Length 380;  
Best Local Similarity 100.0%; Pred. No. 5.7e-09;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. 5831030uhltro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAGUCHI, Kozo  
APPLICANT: YAMAGUCHI, No. 5831030oml  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/472,659  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-067339  
FILING DATE: 04-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-248  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-472-659-3  
Query Match 82.0%; Score 82; DB 2; Length 19;  
Best Local Similarity 84.2%; Pred. No. 1.6e-07;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

CURRENTAPPLICATION DATA:
APPLICATION NUMBER: US/08/568,147B
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 800,952
FILING DATE: 02-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 8425
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX: 230 901 SANUS UR
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 390 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-568-147B-2

Query Match          50.0%; Score 50; DB 1; Length 390;
Best Local Similarity 55.6%; Pred. No. 1.3;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0.

QY      1 VERVDFTHLEDTRRNIN 18
        ||||| |::|:| |
Db       131 VESVDFAANAPESRRKIN 148

RESULT 8
US-08-135-510-5
Sequence 5, Application US/08135510
Patent No. 5420028
GENERAL INFORMATION:
APPLICANT: CHIANG, John Young Ling
TITLE OF INVENTION: Truncated Human Cholesterol
TITLE OF INVENTION: 7 $\alpha$ -Hydroxylase, Method of Production and Use Thereof
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/135,510
FILING DATE: 13-OCT-1993
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: SANDERCOCK, Colin G.
REGISTRATION NUMBER: 31,298
REFERENCE/DOCKET NUMBER: 18748/176 HOCE
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 482 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-135-510-5

Query Match          47.0%; Score 47; DB 1; Length 482;
Best Local Similarity 66.7%; Pred. No. 5.2;
```

APPLICANT: CHIANG, John Young Ling

MEDIUM TYPE: FLIC  
COMPUTER: TRW DC

```

1  TITLE OF INVENTION:  Genomic DNA of Human Cholesterol
2  TITLE OF INVENTION:  7a-Hydroxylase and Methods of Using It
3  NUMBER OF SEQUENCES:  11
4  CORRESPONDENCE ADDRESS:
5  ADDRESSEE:  Foley & Lardner
6  STREET:  3000 K Street, N.W., Suite 500
7  CITY:  Washington
8  STATE:  D.C.
9  COUNTRY:  USA
10 ZIP:  20007-5109
11
12 COMPUTER READABLE FORM:
13 MEDIUM TYPE:  Floppy disk
14
15 COMPUTER:  IBM PC compatible
16 OPERATING SYSTEM:  PC-DOS/MS-DOS
17 SOFTWARE:  Patentin Release #1.0, Version #1.25
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER:  US/08/477,953
20 FILING DATE:  07-JUN-1995
21 PRIOR APPLICATION DATA:
22 APPLICATION NUMBER:  US 08/361,458
23 FILING DATE:  21-DEC-1994
24 PRIOR APPLICATION DATA:
25 APPLICATION NUMBER:  US 08/135,511
26 FILING DATE:  13-OCT-1993
27 PRIOR APPLICATION DATA:
28 APPLICATION NUMBER:  US 08/135,488
29 FILING DATE:  13-OCT-1993
30 PRIOR APPLICATION DATA:
31 APPLICATION NUMBER:  US 08/135,510
32 FILING DATE:  13-OCT-1993
33 ATTORNEY/AGENT INFORMATION:
34 NAME:  SANDERCOCK, Colin G.
35 REGISTRATION NUMBER:  31,298
36 REFERENCE/DOCKET NUMBER:  18748/219 HOCE
37 TELECOMMUNICATION INFORMATION:
38 TELEPHONE:  (202)672-5300
39 TELEFAX:  (202)672-5399
40 TELEX:  904136
41 INFORMATION FOR SEQ ID NO:  5:
42 SEQUENCE CHARACTERISTICS:
43 LENGTH:  482 amino acids
44 TYPE:  amino acid
45 TOPOLOGY:  linear
46
47 US-08-477-953-5
48
49 Query Match      47.0%; Score 47; DB 1; Length 482;
50 Best Local Similarity 66.7%; Pred. No. 5.2;
51 Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0.
52
53 QY      5 DFTNHLDTRRINK 19
54      ||| |||| || |
55 Db      347 DFTLHLEDGSYNIRK 361
56
57 RESULT 11
58 US-08-477-952-5
59 Sequence 5, Application US/08477952
60 Patent No. 5851780
61
62 GENERAL INFORMATION:
63 APPLICANT:  CHIANG, John Young Ling
64 TITLE OF INVENTION:  Genomic DNA of Human Cholesterol
65 TITLE OF INVENTION:  7a-Hydroxylase and Methods of Using It
66 NUMBER OF SEQUENCES:  11
67 CORRESPONDENCE ADDRESS:
68 ADDRESSEE:  Foley & Lardner
69 STREET:  3000 K Street, N.W., Suite 500
70 CITY:  Washington
71 STATE:  D.C.
72 COUNTRY:  USA
73 ZIP:  20007-5109
74
75 COMPUTER READABLE FORM:
76 MEDIUM TYPE:  Floppy disk
77
78 COMPUTER:  IBM PC compatible
79

```

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/477,952  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/361,458  
FILING DATE: 21-DEC-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/135,511  
FILING DATE: 13-OCT-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/135,488  
FILING DATE: 13-OCT-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/135,510  
FILING DATE: 13-OCT-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: SANDERCOCK, Colin G.  
REGISTRATION NUMBER: 31,298  
REFERENCE/DOCKET NUMBER: 18748/221 HOCCE  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 482 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-477-952-5

Query Match 47.0%; Score 47; DB 2; Length 482;  
Best Local Similarity 66.7%; Pred. No. 5.2;  
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 5 DFTNHLEDTRRNINK 19  
Db 347 DFTLHLEDGSYNIRK 361

RESULT 12  
US-08-135-511-29  
Sequence 29, Application US/08135511  
Patent No. 5558999  
GENERAL INFORMATION:  
APPLICANT: Chiang, John  
TITLE OF INVENTION: Cholesterol 7a-Hydroxylase Gene  
TITLE OF INVENTION: Regulatory Elements and Methods for Using Them  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington, D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/135,511  
FILING DATE: 13-OCT-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: SANDERCOCK, Colin G.  
REGISTRATION NUMBER: 31,298  
REFERENCE/DOCKET NUMBER: 18748/175  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136

INFORMATION FOR SEQ ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 503 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-135-511-29

Query Match 47.0%; Score 47; DB 1; Length 503;  
Best Local Similarity 66.7%; Pred. No. 5.5;  
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 5 DFTNHLEDTRRNINK 19  
Db 369 DFTLHLEDGSYNIRK 383

RESULT 13  
US-08-187-453-29  
Sequence 29, Application US/08187453  
Patent No. 5753431  
GENERAL INFORMATION:  
APPLICANT: Chiang, John  
TITLE OF INVENTION: Cholesterol 7a-Hydroxylase Gene  
TITLE OF INVENTION: Regulatory Elements and Transcription Factors  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington, D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/187,453  
FILING DATE: 28-JAN-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/135,488  
FILING DATE: 13-OCT-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/135,511  
FILING DATE: 13-OCT-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/135,510  
FILING DATE: 13-OCT-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: SANDERCOCK, Colin G.  
REGISTRATION NUMBER: 31,298  
REFERENCE/DOCKET NUMBER: 18748/188  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 503 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-187-453-29

Query Match 47.0%; Score 47; DB 1; Length 503;  
Best Local Similarity 66.7%; Pred. No. 5.5;  
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 5 DFTNHLEDTRRNINK 19



DB 369 DFTLHLEDGSYNIRK 383

## RESULT 14

US-08-135-511-28

Sequence 28, Application US/08135511

Patent No. 5558999

GENERAL INFORMATION:

APPLICANT: Chiang, John

TITLE OF INVENTION: Cholesterol 7a-Hydroxylase Gene

TITLE OF INVENTION: Regulatory Elements and Methods for Using Them

NUMBER OF SEQUENCES: 35

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner

STREET: 3000 K Street, N.W., Suite 500

CITY: Washington, D.C.

COUNTRY: USA

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08-135,511

FILING DATE: 13-OCT-1993

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: SANDERCOCK, Colin G.

REGISTRATION NUMBER: 31,298

REFERENCE/DOCKET NUMBER: 18748/175

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202)672-5300

TELEFAX: (202)672-5399

TELEX: 904136

INFORMATION FOR SEQ ID NO: 28:

SEQUENCE CHARACTERISTICS:

LENGTH: 504 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-135-511-28

Query Match 47.0%; Score 47; DB 1; Length 504;  
Best Local Similarity 66.7%; Pred. No. 5.5;  
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 5 DFTNHLDTRRNIRK 19

DB 369 DFTLHLEDGSYNIRK 383

## RESULT 15

US-08-135-511-30

Sequence 30, Application US/08135511

Patent No. 5558999

GENERAL INFORMATION:

APPLICANT: Chiang, John

TITLE OF INVENTION: Cholesterol 7a-Hydroxylase Gene

TITLE OF INVENTION: Regulatory Elements and Methods for Using Them

NUMBER OF SEQUENCES: 35

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner

STREET: 3000 K Street, N.W., Suite 500

CITY: Washington, D.C.

COUNTRY: USA

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/135,511

FILING DATE: 13-OCT-1993

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: SANDERCOCK, Colin G.

REGISTRATION NUMBER: 31,298

REFERENCE/DOCKET NUMBER: 18748/175

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202)672-5300

TELEFAX: (202)672-5399

TELEX: 904136

INFORMATION FOR SEQ ID NO: 30:

SEQUENCE CHARACTERISTICS:

LENGTH: 504 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-135-511-30

Query Match 47.0%; Score 47; DB 1; Length 504;  
Best Local Similarity 66.7%; Pred. No. 5.5;  
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 5 DFTNHLDTRRNIRK 19

DB 369 DFTLHLEDGSYNIRK 383

Search completed: July 11, 2003, 12:02:06  
Job time : 4.39709 secs

**THIS PAGE BLANK (USPIO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:59:17 ; Search time 5.56965 Seconds  
(Without alignments)  
397.148 Million cell updates/sec

Title: US-10-091-442-3  
Perfect score: 100  
Sequence: 1 VERVDFTNHLEDRNRINK 19

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 445758 seqs, 116419773 residues

Total number of hits satisfying chosen parameters: 445758

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications\_AA:\*  
1: /cgn2\_6/ptodata/1/pubppaa/US07\_NEW\_PUB.pep:\*  
2: /cgn2\_6/ptodata/1/pubppaa/PC7\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubppaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubppaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubppaa/PC7US\_PUBCOMB.pep:\*  
6: /cgn2\_6/ptodata/1/pubppaa/US07\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubppaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubppaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubppaa/US09\_NEW\_PUB.pep:\*  
10: /cgn2\_6/ptodata/1/pubppaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubppaa/US10\_NEW\_PUB.pep:\*  
12: /cgn2\_6/ptodata/1/pubppaa/US10\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/1/pubppaa/US60\_NEW\_PUB.pep:\*  
14: /cgn2\_6/ptodata/1/pubppaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	100	100.0	19	US-10-091-442-3	Sequence 3, Appl1
2	100	100.0	19	US-09-140-719-3	Sequence 34, Appl1
3	100	100.0	380	US-10-091-442-34	Sequence 34, Appl1
4	100	100.0	380	US-09-140-719-34	Sequence 34, Appl1
5	50	50.0	390	US-08-731-566-2	Sequence 29, Appl1
6	50	50.0	390	US-09-823-187-29	Sequence 29, Appl1
7	50	50.0	390	US-09-823-187-30	Sequence 27, Appl1
8	46	46.0	390	US-09-823-187-27	Sequence 27, Appl1
9	46	46.0	390	US-09-823-187-28	Sequence 32, Appl1
10	46	46.0	410	US-09-910-186A-32	Sequence 32, Appl1
11	45	45.0	45	US-09-864-761-41911	Sequence 41911, A
12	45	45.0	123	US-10-073-979-21	Sequence 21, Appl1
13	45	45.0	123	US-09-764-891-4883	Sequence 4883, Ap
14	45	45.0	123	US-09-764-898-242	Sequence 242, App
15	45	45.0	391	US-10-157-031-355	Sequence 355, App
16	45	45.0	391	US-09-466-386A-110	Sequence 110, App
17	45	45.0	391	US-09-735-705-110	Sequence 110, App
18	45	45.0	391	US-09-850-716A-110	Sequence 110, App
19	45	45.0	391	US-09-897-778-110	Sequence 110, App

20	45	45.0	400	US-09-466-396A-112	Sequence 112, App
21	45	45.0	400	US-09-735-705-112	Sequence 112, App
22	45	45.0	400	US-09-850-716A-112	Sequence 112, App
23	45	45.0	400	US-09-897-778-112	Sequence 112, App
24	45	45.0	612	US-09-925-301-1358	Sequence 1358, Ap
25	44	44.0	340	US-09-764-898-169	Sequence 169, App
26	44	44.0	392	US-10-198-070-62	Sequence 62, Appl
27	44	44.0	617	US-10-012-696-947	Sequence 947, App
28	44	44.0	617	US-09-895-814-947	Sequence 947, App
29	42	42.0	379	US-10-012-896-946	Sequence 946, App
30	42	42.0	379	US-09-885-814-946	Sequence 946, App
31	42	42.0	392	US-10-198-070-67	Sequence 67, Appl
32	42	42.0	392	US-10-198-070-70	Sequence 70, Appl
33	41	41.0	395	US-09-823-187-2	Sequence 2, Appl1
34	41	41.0	437	US-09-925-301-1200	Sequence 1200, Ap
35	41	41.0	459	US-09-925-300-1440	Sequence 1440, Ap
36	41	41.0	531	US-10-156-761-14639	Sequence 14639, A
37	41	41.0	558	US-09-941-831-28	Sequence 28, Appl
38	41	41.0	705	US-10-128-714-3201	Sequence 3201, Ap
39	41	41.0	748	US-10-128-714-8201	Sequence 8201, Ap
40	41	41.0	818	US-09-738-626-6491	Sequence 6491, Ap
41	40.5	40.5	1019	US-09-801-574-76	Sequence 76, Appl
42	40	40.0	150	US-10-001-857-203	Sequence 203, Appl
43	40	40.0	433	US-09-738-626-6592	Sequence 6592, Ap
44	39.5	39.5	745	US-10-000-823-4	Sequence 4, Appl1
45	39.5	39.5	761	US-10-011-582-10	Sequence 10, Appl

#### ALIGNMENTS

RESULT 1  
US-10-091-442-3  
Sequence 3, Application US/10091442  
; Patent No. US20020164711A1  
; GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
IWASA, Fuyuki  
TSURUOKA, No. US20020164711A1uo  
NAKAZATO, Hiroshi  
MIDURA, Kenju  
ISHIDA, No. US20020164711A1uhiro  
KURIHARA, Tatsuya  
YAMACHI, Kozo  
YAMAGUCHI, No. US20020164711A1omi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESSES:  
ADDRESS: Burns, Doane, Swecker & Mathis  
STREET: P. O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/091.442  
FILING DATE: 07-Mar-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/140.719  
FILING DATE: 08-AUG-1998  
APPLICATION NUMBER: US 08/474.661  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/091.028  
FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
APPLICATION NUMBER: JP 6-067339

FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-10-091-442-3

Query Match 100.0%; Score 100; DB 9; Length 19;  
Best Local Similarity 100.0%; Pred. No. 6.8e-10;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VERDYFTNHLEDRNRINK 19  
DB 1 VERDYFTNHLEDRNRINK 19

RESULT 2  
US-09-140-719-3  
Sequence 3, Application US/09140719  
Patent No. US20010026931A1  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSUROJOKA, No. US20010026931A1uo  
APPLICANT: MAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. US20010026931A1uhiro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, No. US20010026931A1omi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
City: Alexandria  
State: Virginia  
Country: United States  
Zip: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.

REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-140-719-3

Query Match 100.0%; Score 100; DB 10; Length 19;  
Best Local Similarity 100.0%; Pred. No. 6.8e-10;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VERDYFTNHLEDRNRINK 19  
DB 1 VERDYFTNHLEDRNRINK 19

RESULT 3  
US-10-091-442-34  
Sequence 34, Application US/10091442  
Patent No. US20020164711A1  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASA, Fuyuki  
APPLICANT: TSUROJOKA, No. US20020164711A1uo  
APPLICANT: MAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. US20020164711A1uhiro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAICHI, Kozo  
APPLICANT: YAMAGUCHI, No. US20020164711A1omi  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
City: Alexandria  
State: Virginia  
Country: United States  
Zip: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/091,442  
FILING DATE: 07-Mar-2002.  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021

INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 34:  
US-10-091-442-34

Query Match 100.0%; Score 100; DB 9; Length 380;  
Best Local Similarity 100.0%; Pred. No. 2,1e-08;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VERVDFTNHLNEDTRNIN 19  
DB 126 VERVDFTNHLNEDTRNIN 144

RESULT 4  
US-09-140-719-34  
Sequence 34, Application US/09140719  
Patent No. US20010026931A1  
GENERAL INFORMATION:  
APPLICANT: TSUJIMOTO, Masafumi  
APPLICANT: IWASHI, Fuyuki  
APPLICANT: TSURUOKA, No. US20010026931A1uo  
APPLICANT: NAKAZATO, Hiroshi  
APPLICANT: MIURA, Kenju  
APPLICANT: ISHIDA, No. US20010026931A1uhlro  
APPLICANT: KURIHARA, Tatsuya  
APPLICANT: YAMAGUCHI, Kozo  
TITLE OF INVENTION: MEGAKARYOCYTE DIFFERENTIATION FACTOR  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/140,719  
FILING DATE: 08-AUG-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/474,661  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/091,028  
FILING DATE: 14-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-212305  
FILING DATE: 17-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-067339  
FILING DATE: 04-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 001560-247  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 380 amino acids  
TYPE: amino acid

TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-140-719-34

Query Match 100.0%; Score 100; DB 10; Length 380;  
Best Local Similarity 100.0%; Pred. No. 2,1e-08;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VERVDFTNHLNEDTRNIN 19  
DB 126 VERVDFTNHLNEDTRNIN 144

RESULT 5  
US-08-731-566-2  
Sequence 2, Application US/08731566  
Publication No. US20030072752A1  
GENERAL INFORMATION:  
APPLICANT: Toshihiko Hibino et al.  
TITLE OF INVENTION: TWO NOVEL GENES FROM PSORIATIC EPIDERMIS:  
PSORIASIS TYPE I AND PSORIASIS TYPE II  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD, LLP  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1775  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/731,566  
FILING DATE: 16-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/013,755  
FILING DATE: 20-MAR-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/005,679  
FILING DATE: 17-OCT-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Louis Myers  
REGISTRATION NUMBER: 35,965  
REFERENCE/DOCKET NUMBER: MGP-040CP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)227-5941  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 390 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-731-566-2

Query Match 50.0%; Score 50; DB 7; Length 390;  
Best Local Similarity 55.6%; Pred. No. 3;  
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 1 VERVDFTNHLNEDTRNIN 18  
DB 131 VESYDFANAPESRRKIN 148

RESULT 6  
US-09-823-187-29  
Sequence 29, Application US/09823187  
Publication No. US20030096952A1  
GENERAL INFORMATION:  
APPLICANT: Burgess, Catherine

```

Query Match          50.0%; Score 50; DB 9; Length 390;
Best Local Similarity 55.6%; Pred. No. 3;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0

OY      1 VERVDFTNHEDTRRNIN 18
      ||||| 1::1: ||
Db      131 VESYDFANAPESRRKIN 148

RESULT 7
US-09-823-187-30
: Sequence 30, Application US/09823187
: Publication No. US2003009652A1
: GENERAL INFORMATION:
: APPLICANT: Burgess, Catherine
: APPLICANT: Gusev, Vladimir Y
: APPLICANT: Liu, Xiaohong
: APPLICANT: Majumder, Kunud
: APPLICANT: Padigaru, Muralidhar
: APPLICANT: Patturajan, Meera
: APPLICANT: Shinkets, Richard A
: APPLICANT: Spaderna, Steven K
: APPLICANT: Spytek, Kimberly
: APPLICANT: Taulpie, Raymond J
: TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
: FILE REFERENCE: 15966-745
: CURRENT APPLICATION NUMBER: US/09/823,187
: CURRENT FILING DATE: 2001-03-29
: PRIOR APPLICATION NUMBER: 60/193,339
: PRIOR FILING DATE: 2000-03-30
: PRIOR APPLICATION NUMBER: 60/193,205
: PRIOR FILING DATE: 2000-03-30

```

```

: PRIOR APPLICATION NUMBER: 60/195,343
: PRIOR FILING DATE: 2000-04-05
: PRIOR APPLICATION NUMBER: 60/195,088
: PRIOR FILING DATE: 2000-04-06
: PRIOR APPLICATION NUMBER: 60/195,005
: PRIOR FILING DATE: 2000-04-06
: PRIOR APPLICATION NUMBER: 60/195,792
: PRIOR FILING DATE: 2000-04-10
: PRIOR APPLICATION NUMBER: 60/196,556
: PRIOR FILING DATE: 2000-04-11
: PRIOR APPLICATION NUMBER: 60/197,081
: PRIOR FILING DATE: 2000-04-13
: PRIOR APPLICATION NUMBER: 60/197,525
: PRIOR FILING DATE: 2000-04-14
: PRIOR APPLICATION NUMBER: 60/197,087
: PRIOR FILING DATE: 2000-04-14
: NUMBER OF SEQ ID NOS: 103
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 30
: LENGTH: 390
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-823-187-30

Query Match          50.0%; Score 50; DB 9; Length 390;
Best Local Similarity 55.6%; Pred. No. 3;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0
Oy      1 VERVDFTHLETRRNN 18
      ||||| |::|: ||
Db      131 VESYDFANAPESRRKKN 148

RESULT 8
US-09-823-187-27
: Sequence 27, Application US/09823187
: Publication NO. US20030096952A1
: GENERAL INFORMATION:
: APPLICANT: Burgess, Catherine
: APPLICANT: Gusev, Vladimir Y
: APPLICANT: Liu, Xiaohong
: APPLICANT: Majumder, Kundu
: APPLICANT: Padigar, Muralidhar
: APPLICANT: Patursajan, Meera
: APPLICANT: Shimkels, Richard A
: APPLICANT: Spaderna, Steven K
: APPLICANT: Spytek, Kimberly
: APPLICANT: Taupier, Raymond J
: TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
: FILE REFERENCE: 15966-745
: CURRENT APPLICATION NUMBER: US/09/823,187
: CURRENT FILING DATE: 2001-03-29
: PRIOR APPLICATION NUMBER: 60/193,339
: PRIOR FILING DATE: 2000-03-30
: PRIOR APPLICATION NUMBER: 60/193,205
: PRIOR FILING DATE: 2000-03-30
: PRIOR APPLICATION NUMBER: 60/195,343
: PRIOR FILING DATE: 2000-04-05
: PRIOR APPLICATION NUMBER: 60/195,088
: PRIOR FILING DATE: 2000-04-06
: PRIOR APPLICATION NUMBER: 60/195,005
: PRIOR FILING DATE: 2000-04-06
: PRIOR APPLICATION NUMBER: 60/195,792
: PRIOR FILING DATE: 2000-04-10
: PRIOR APPLICATION NUMBER: 60/196,556
: PRIOR FILING DATE: 2000-04-11
: PRIOR APPLICATION NUMBER: 60/197,081
: PRIOR FILING DATE: 2000-04-13
: PRIOR APPLICATION NUMBER: 60/197,525
: PRIOR FILING DATE: 2000-04-14
: PRIOR APPLICATION NUMBER: 60/197,087
: PRIOR FILING DATE: 2000-04-14
: NUMBER OF SEQ ID NOS: 103

```

## RESULT 10

; PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 41911  
LENGTH: 45  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC023009.3  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.2  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.3  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.1  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.2  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.2  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.1  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.8  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.1  
OTHER INFORMATION: SWISSPROT HIT: P00488, EVALUATE 8.00e-21  
OTHER INFORMATION: EST\_HUMAN HIT: AV733491.1, EVALUATE 3.00e-15  
US-09-864-761-41911

Query Match 45.0%; Score 45; DB 10; Length 45;  
Best Local Similarity 57.1%; Pred. No. 1.7;  
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 4 VDFTNHLEDTNRNI 17  
DB 15 VEFTNPLKETLRNV 28

RESULT 12  
US-10-073-979-21  
Sequence 21, Application US/10073979  
Publication No. US20030044905A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PT231C1  
CURRENT APPLICATION NUMBER: US/10/073,979  
CURRENT FILING DATE: 2002-02-14  
Prior Application removed - See file Wrapper or Palm  
NUMBER OF SEQ ID NOS: 22  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 21  
LENGTH: 123  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (88)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-10-073-979-21

Query Match 45.0%; Score 45; DB 9; Length 123;  
Best Local Similarity 44.4%; Pred. No. 5.3;  
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 VERVDFTNHLEDTNRNIN 18  
DB 82 LQTVDFXQSTETFRKTN 99

RESULT 13  
US-09-764-891-4883  
Sequence 4883, Application US/09764891  
Publication No. US20030077808A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PC006  
CURRENT APPLICATION NUMBER: US/09/764,891  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 10231  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 4883  
LENGTH: 123  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (88)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-891-4883

Query Match 45.0%; Score 45; DB 9; Length 123;  
Best Local Similarity 44.4%; Pred. No. 5.3;  
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 VERVDFTNHLEDTNRNIN 18  
DB 82 LQTVDFXQSTETFRKTN 99

RESULT 14  
US-09-764-898-242  
Sequence 242, Application US/09764898  
Patent No. US20020090673A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PJ201  
CURRENT APPLICATION NUMBER: US/09/764,898  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 311  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 242  
LENGTH: 123  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (88)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-898-242

Query Match 45.0%; Score 45; DB 10; Length 123;  
Best Local Similarity 44.4%; Pred. No. 5.3;  
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 VERVDFTNHLEDTNRNIN 18  
DB 82 LQTVDFXQSTETFRKTN 99



Db 82 LQVDFXQSTTEKRTIN 99

RESULT 15

US-10-157-031-355

; Sequence 355, Application US/10157031  
; Publication No. US20030108890A1

; GENERAL INFORMATION:

; APPLICANT: Baranova, A. V.

; APPLICANT: Yankovsky, N. K.

; APPLICANT: Kozlov, A. P.

; APPLICANT: Lodashev, A. V.

; APPLICANT: Krukovskaya, L. L.

; TITLE OF INVENTION: In silico screening for phenotype-associated expressed sequences

; FILE REFERENCE: 2760-103

; CURRENT APPLICATION NUMBER: US/10/157,031

; CURRENT FILING DATE: 2002-05-30

; NUMBER OF SEQ ID NOS: 415

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 355

; LENGTH: 391

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-157-031-355

Query Match 45.0%; Score 45; DB 9; Length 391;

Best Local Similarity 44.4%; Pred. No. 20;

Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

OY 1 VERVDFTHLEDTRRIN 18

Db 132 LQVDFVNADESRKKIN 149

Search completed: July 11, 2003, 12:37:32  
Job time : 5.56965 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: July 11, 2003, 11:45:06 ; Search time 3.8316 Seconds  
(without alignments)  
476.708 Million cell updates/sec

Title: US-10-091-442-3

Perfect score: 100

Sequence: 1 VERVDFTNHLEDFRRNINK.19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: PIR\_73:\*  
2: PIR1:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	50.0	390	2	I38201
2	49	49.0	1169	2	H70178
3	48	48.0	472	2	T27755
4	47	47.0	501	2	I46701
5	47	47.0	503	1	A35376
6	47	47.0	503	2	A54779
7	47	47.0	504	1	JH0659
8	47	47.0	504	2	S39399
9	47	47.0	1046	2	T42720
10	47	47.0	1046	2	T42734
11	46	46.0	375	2	A36898
12	46	46.0	390	2	I38202
13	46	46.0	1287	2	S39791
14	46	46.0	2022	2	T43214
15	45.5	45.5	163	2	D64427
16	45	45.0	391	2	JC7118
17	45	45.0	673	1	B64530
18	45	45.0	673	2	A71978
19	44.5	44.5	330	2	D97235
20	44.5	44.5	333	2	S57050
21	44	44.0	348	2	E82933
22	44	44.0	369	2	T48612
23	44	44.0	483	2	JC5743
24	44	44.0	555	1	SYHOMA
25	44	44.0	1147	2	S64930
26	43.5	43.5	119	2	S76276
27	43	43.0	58	2	G95132
28	43	43.0	259	2	G81833
29	43	43.0	301	2	C64633

30	43	43.0	363	2	F71881	probable flagellar
31	43	43.0	468	2	F84686	hypothetical prote
32	43	43.0	732	1	EKHUX	protein-glutamine
33	43	43.0	1111	2	A86922	probable arabinosy
34	43	43.0	1461	2	T26327	hypothetical prote
35	43	43.0	1619	2	T18499	hypothetical prote
36	42.5	42.5	409	2	A87495	hypothetical prote
37	42	42.0	399	2	E71188	hypothetical prote
38	42	42.0	589	2	S71954	RNA/ssDNA-binding
39	42	42.0	602	2	S23991	conjugative transf
40	42	42.0	957	2	H69141	hypothetical prote
41	41.5	41.5	349	2	AH1764	B. subtilis probab
42	41	41.0	288	2	AD0884	probable transfera
43	41	41.0	317	2	T27330	hypothetical prote
44	41	41.0	319	2	F96930	molybdate-binding
45	41	41.0	347	2	S44846	K06H7.6 protein -

#### ALIGNMENTS

RESULT 1  
I38201  
squamous cell carcinoma antigen.1 - human  
C:Species: Homo sapiens (man)  
C>Date: 23-Feb-1996 #sequence.revision 23-Feb-1996 #text\_change 26-May-2000  
C:Accession: I38201; I38200; G01631; J70966; J70967  
R:Schneider, S.S.; Schick, C.; Fish, K.E.; Miller, E.; Pena, J.C.; Treter, S.D.; Hul,  
Proc. Natl. Acad. Sci. U.S.A. 92, 3147-3151, 1995  
A>Title: A serine proteinase inhibitor locus at 16q21.3 contains a tandem duplication  
A:Reference number: I38200; M0ID:95241462; PMID:7724531  
A:Accession: I38201  
A>Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-390 <SCH1>  
A:Cross-references: EMBL:U19568; GB:U19558; NID:q1172085  
A:Accession: I38200  
A>Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-117 <SCH2>  
A:Cross-references: EMBL:U19562; NID:q897835; PIDN:AAH8516.1; PID:q897844  
R:Silverman, G.A.  
submitted to the EMBL Data Library, January 1995  
A:Reference number: G07968  
A:Accession: G01631  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-350, 'G', 352-390 <SHL>  
A:Cross-references: EMBL:U19556; NID:q1276435; PID:q1052869  
R:Sumitani, Y.; Kishi, F.; Sekiguchi, K.; Kato, H.  
Biochem. Biophys. Res. Commun. 181, 51-58, 1991  
A>Title: Squamous cell carcinoma antigen is a new member of the serine protease inhib  
A:Reference number: J70966; M0ID:92068241; PMID:1958219  
A:Accession: J70966  
A:Molecule type: mRNA  
A:Residues: 1-350, 'G', 352-356, 'A', 358-390 <SUM1>  
A:Cross-references: GB:S66896; NID:q239551; PIDN:AAH20405.1; PID:q239552  
A:Accession: J70967  
A:Molecule type: protein  
A:Residues: 11-21; 231-237; 240-256; 303-325 <SUM2>  
C:Comment: This antigen probably acts as a proteinase inhibitor to modulate the host  
C:Genetics:  
A:Gene: GDB:SCCA1; SCC  
A:Cross-references: GDB:625364; OMIM:600517  
A:Map position: 18q21.3-18q21.3  
A:Introns: 55/3; 74/3; 117/3; 157/1; 204/3; 256/3  
C:Superfamily: antithrombin III  
C:Keywords: cysteine proteinase inhibitor; glycoprotein  
F:65,93,171,376/binding site: carboxylate (Asn) (covalent) #status predicted  
F:334/inhibitory site: Ser (cathepsin L) #status predicted

Query Match 50.0%; Score 50; DB 2; Length 390;  
Best Local Similarity 55.6%; Pred. No. 3.3;

A: Title: structural analysis of the  
U: Title: 200, 040/-04/1, 15  
C: Title: 200, 040/-04/1, 15

A:Reference number: A38736; MUID:91177904; PMID:2007596  
A:Accession: A38736  
A:Molecule type: DNA  
A:Residues: 1-370, 'S', 372-503 <NIS>  
A:CROSS-references: GB:M59184  
R:Chiang, J.Y.L.; Yang, T.P.; Wang, D.P.  
Biochem. Biophys. Acta 1132, 337-339, 1992  
A:Title: Cloning and 5'-flanking sequence of a rat cholesterol 7alpha-hydroxylase gene.  
A:Reference number: S27206; MUID:93041942; PMID:1420318  
A:Accession: S27206  
A:Molecule type: DNA  
A:Residues: 1-26 <CH>  
A:CROSS-references: GB:214108; GB:S48135; NID:955835; PIDN:CAA78481.1; PID:955836  
C:Comment: This liver microsomal enzyme catalyzes the conversion of cholesterol to bile  
C:Genetics:  
A:Introns: 27/2; 107/3; 303/2; 347/1; 405/3  
C:Superfamily: human cytochrome P450 CYP7A1; cytochrome P450 homology  
C:Keywords: cholesterol metabolism; chromoprotein; electron transfer; heme; iron; metal  
F:282-466/Domain: cytochrome P450 homology <P45>  
F:444/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 47.0%; Score 47; DB 1; Length 503;  
Best Local Similarity 66.7%; Pred. No. 13;  
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 5 DFTNHLDETRNIRK 19  
||| ||| ||| |||  
Db 369 DFTLHLEDGSYNIRK 383

RESULT 6  
A54779  
Cholesterol 7alpha-monooxygenase (EC 1.14.13.17) - mouse  
C:Species: Mus musculus (house mouse)  
C>Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 04-Mar-2000  
C:Accession: A54779  
R:Tung, K.W.; Ishimura-Oka, K.; Kihara, S.; Oka, K.; Chan, L.  
Genomics 21, 244-247, 1994  
A:Title: Structure of the mouse cholesterol 7alpha-hydroxylase gene.  
A:Reference number: A54779; MUID:94375022; PMID:8088795  
A:Accession: A54779  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-503 <TZU>  
A:CROSS-references: GB:L23754  
C:Genetics:  
A:Gene: CYP7  
C:Superfamily: human cytochrome P450 CYP7A1; cytochrome P450 homology  
C:Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase; transmembrane pro  
F:282-466/Domain: cytochrome P450 homology <P45>  
F:444/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 47.0%; Score 47; DB 2; Length 503;  
Best Local Similarity 66.7%; Pred. No. 13;  
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 5 DFTNHLDETRNIRK 19  
||| ||| ||| |||  
Db 369 DFTLHLEDGSYNIRK 383

RESULT 7  
JH0659  
cholesterol 7alpha-monooxygenase (EC 1.14.13.17) chain 1 - human  
N:Alternate names: cholesterol 7alpha-hydroxylase; cytochrome P450, subfamily VIIA  
C:Species: Homo sapiens (man)  
C>Date: 30-Jun-1993 #sequence\_revision 30-Jun-1993 #text\_change 03-Mar-2000  
C:Accession: S29818; JH0659; S11051; A42201; A54310  
R:Nishimoto, M.; Noshiro, M.; Okuda, K.  
Biochim. Biophys. Acta 1172, 147-150, 1993  
A:Title: Structure of the gene encoding human liver cholesterol 7alpha-hydroxylase.  
A:Reference number: S29818; MUID:93176797; PMID:8439551  
A:Accession: S29818

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-504 <NTS>  
A:CROSS-references: EMBL:L04633  
R:Karam, W.G.; Chiang, J.Y.L.  
Biochem. Biophys. Res. Commun. 185, 568-595, 1992  
A:Title: Polymorphisms of human cholesterol 7 alpha-hydroxylase.  
A:Reference number: JH0659; MUID:92304280; PMID:1610352  
A:Accession: JH0659  
A:Molecule type: mRNA  
A:Residues: 1-504 <KAR>  
A:CROSS-references: GB:M93133; NID:9181318; PIDN:AAA58435.1; PID:9181319  
A:Experimental source: liver  
A>Note: 100-ser was also found  
R:Noshiro, M.; Okuda, K.  
FEBS Lett. 268, 137-140, 1990  
A:Title: Molecular cloning and sequence analysis of cDNA encoding human cholesterol 7  
A:Reference number: S11051; MUID:90346120; PMID:2384150  
A:Accession: S11051  
A:Molecule type: mRNA  
A:Residues: 1-346, 'N', 348-384, 'S', 386-504 <NOS>  
A:CROSS-references: GB:M93133  
R:Molowa, D.T.; Chen, W.S.; Cima, G.M.; Tan, C.P.  
Biochemistry 31, 2539-2544, 1992  
A:Title: Transcriptional regulation of the human cholesterol 7 alpha-hydroxylase gene  
A:Reference number: A42201; MUID:92190183; PMID:1312351  
A:Accession: A42201  
A:Molecule type: DNA  
A:Residues: 1-25 <MOL>  
A:CROSS-references: GB:M9647; GB:J05363; NID:9180469; PIDN:AAA58423.1; PID:9553228  
A>Note: Sequence extracted from NCBI backbone (NCBI:89078, NCBI:89079)  
R:Wang, D.P.; Chiang, J.Y.  
Genomics 20, 320-323, 1994  
A:Title: Structure and nucleotide sequences of the human cholesterol 7 alpha-hydroxyl  
A:Reference number: A54310; MUID:94292222; PMID:8020987  
A:Accession: A54310  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-140 <RES>  
A:CROSS-references: GB:L13460; NID:949885; PIDN:AAA61350.1; PID:9624966  
C:Comment: This liver microsomal enzyme catalyzes the conversion of cholesterol to bl  
C:Genetics:  
A:Gene: GDB:CYP7A1; CYP7  
A:CROSS-references: GDB:132221; OMIM:118455  
A:Map position: 8q11-8q12  
A:Introns: 27/2; 107/3; 303/2; 347/1; 405/3  
C:Superfamily: human cytochrome P450 CYP7A1; cytochrome P450 homology  
C:Keywords: cholesterol metabolism; chromoprotein; electron transfer; heme; iron; met  
F:282-466/Domain: cytochrome P450 homology <CYP>  
F:444/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 47.0%; Score 47; DB 1; Length 504;  
Best Local Similarity 66.7%; Pred. No. 13;  
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 5 DFTNHLDETRNIRK 19  
||| ||| ||| |||  
Db 369 DFTLHLEDGSYNIRK 383

RESULT 8  
S3939  
cholesterol 7alpha-monooxygenase (EC 1.14.13.17) - Chinese hamster  
N:Alternate names: cholesterol 7alpha-hydroxylase  
C:Species: Cricetus griseus (Chinese hamster)  
C>Date: 07-Oct-1994 #sequence\_revision 01-Dec-1995 #text\_change 04-Mar-2000  
C:Accession: S3939  
R:Crestani, M.; Galli, G.; Chiang, J.Y.L.  
Arch. Biochem. Biophys. 306, 451-460, 1993  
A:Title: Genomic cloning, sequencing, and analysis of the hamster cholesterol 7-alpha  
A:Reference number: S3939; MUID:94029018; PMID:8105753  
A:Accession: S3939  
A:Status: preliminary

Query Match	47.0%;	Score 47;	DB 2;	Length 1046;
-------------	--------	-----------	-------	--------------

C;Keywords: glycoprotein; serine proteinase inhibitor  
E:1-32/Domain: signal sequence #status predicted <STG>  
C;Superfamily: antitrypsin III



**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 11:43:41 ; Search time 1.81705 Seconds  
(without alignments)

433.698 Million cell updates/sec

Title: US-10-091-442-3

Perfect score: 100

Sequence: 1 VERVDFTHLEDTERRINK 19

Scoring table: BLOSUM62

Gapped 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

Result No. Score Query Match Length DB ID Description

Result No.	Score	Query Match	Length	DB ID	Description
1	100	100.0	380	1 SPB7_HUMAN	O75635 homo sapien
2	50	50.0	390	1 SCC1_HUMAN	P29508 homo sapien
3	49	49.0	1169	1 EX5B_BORBU	O51578 borrellia bu
4	48.5	48.5	210	1 NOC4_HUMAN	O43402 homo sapien
5	47.5	47.5	207	1 NOC4_MOUSE	O70378 mus musculi
6	47	47.0	501	1 CP7A_MOUSE	P51342 oryctolagus
7	47	47.0	503	1 CP7A_RABIT	O64505 mus musculi
8	47	47.0	503	1 CP7A_RAT	P18125 rattus norv
9	47	47.0	504	1 CP7A_CRIGR	P46634 cricetulus
10	47	47.0	504	1 CP7A_HUMAN	P22680 homo sapien
11	46	46.0	375	1 MASP_HUMAN	P36352 homo sapien
12	46	46.0	375	1 MASP_MOUSE	P70124 mus musculi
13	46	46.0	375	1 MASP_RAT	P70564 rattus norv
14	46	46.0	390	1 SCC2_HUMAN	O60393 clostridium
15	46	46.0	1296	1 BNG_CLOBO	O58427 methanococ
16	46	46.0	2022	1 ANTL_ONCYO	O58427 methanococ
17	45.5	45.5	163	1 YAL1_METUA	O94108 homo sapien
18	45	45.0	391	1 SBI3_HUMAN	O94108 homo sapien
19	44.5	44.5	483	1 CYP7_YEAST	O96815 homo sapien
20	44	44.0	501	1 CP7A_PIG	O96815 homo sapien
21	44	44.0	501	1 CP7A_PIG	O96815 homo sapien
22	44	44.0	555	1 MASY_PICAN	P21360 pichia angu
23	44	44.0	1147	1 KIN2_YEAST	P13186 saccharomy
24	43	43.0	505	1 CSC1_DROME	O94108 homo sapien
25	43	43.0	731	1 F13A_HUMAN	O94108 homo sapien
26	43	43.0	1111	1 EMBA_MYCLE	O96815 homo sapien
27	42	42.0	392	1 SBI1_HUMAN	O96815 homo sapien
28	42	42.0	592	1 RB56_HUMAN	O96815 homo sapien
29	42	42.0	602	1 TRAN_ECOTI	P24082 escherichia
30	41	41.0	203	1 U172_DROME	O94108 homo sapien
31	41	41.0	215	1 HMUO_CORDI	P71119 corbynebacte
32	41	41.0	285	1 STAR_BRARE	O94108 homo sapien
33	41	41.0	376	1 PT16_HUMAN	P35237 homo sapien

34	41	41.0	378	1 PT16_BOVIN	O02739 bos taurus
35	41	41.0	379	1 ILEU_HUMAN	P30740 homo sapien
36	41	41.0	405	1 SBI2_HUMAN	O96815 homo sapien
37	41	41.0	728	1 YMR6_CAEEL	P34514 caenorhabdi
38	41	41.0	1161	1 DP3A_EUCAL	P57332 buchemera ap
39	41	41.0	1195	1 YCX7_CHLRE	P36495 chlamydomon
40	41	41.0	2364	1 SPCO_HUMAN	O01082 homo sapien
41	40.5	40.5	578	1 YDEM_CAEEL	O19124 caenorhabdi
42	40.5	40.5	842	1 SYA_CAMEL	O9p105 campylobact
43	40.5	40.5	1019	1 STJ1_HUMAN	O9p105 homo sapien
44	40	40.0	171	1 VEP2_BPP2	P22502 bacterioph
45	40	40.0	289	1 STX3_HUMAN	O13277 homo sapien

## ALIGNMENTS

RESULT 1  
ID SPB7\_HUMAN STANDARD: PRT; 380 AA.  
AC O75635:  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Megsin (TP55) (Serpin B7).  
GN SERPINB7.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97326116; PubMed=9182567;  
RA Tsumoto M., Tsumoto N., Ishida N., Kurihara T., Iwasa F.,  
RA Yamashiro K., Rogi T., Kodama S., Katsuragi N., Adachi M.,  
RA Katayama T., Nakao M., Yamachi K., Hashino J., Haruyama M., Miura K.,  
RA Nakanishi T., Nakazato H., Teramura M., Mizoguchi H., Yamaguchi N.,  
RT "Purification, cDNA cloning, and characterization of a new serpin with  
RT megakaryocyte maturation activity".  
RL J. Biol. Chem. 272:15373-15380(1997).  
[2]  
RN SEQUENCE FROM N.A.  
RP TISSUE=Mesangial cells;  
RX MEDLINE=98376492; PubMed=9710452;  
RA Miyata T., Nangaku M., Suzuki D., Inagi R., Uregami K., Sakai H.,  
RA Okubo K., Kurokawa K.;  
RT "A mesangium-predominant gene, megisin, is a new serpin upregulated in  
RT Iga nephropathy".  
RL J. Clin. Invest. 102:828-836(1998).  
CC -!- FUNCTION: Might function as an inhibitor of lys-specific  
CC proteases. Might influence the maturation of megakaryocytes via  
CC its action as a serpin.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -!- TISSUE SPECIFICITY: Predominantly expressed in mesangial cells.  
CC -!- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL: D88575; BAA31232.1; -  
CC EMBL: AF027866; AAC64506.1; -  
CC HSSP: P05619; 1HLE.  
CC Gene: HGNC:13902; SERPINB7.  
CC MIM: 603357; -  
CC InterPro: IPR000215; Serpin.  
CC Pfam: PF00079; serpin; 1.  
CC SMART: SM00093; SERPIN; 1.  
CC PROSITE: PS00284; SERPIN; 1.

KW Serpin; Serine protease inhibitor.  
 FT ACT\_SITE 347 348 REACTIVE BOND (BY SIMILARITY).  
 SQ SEQUENCE 380 AA; 42904 MW; 9A2CDB6C63CFF605 CRC64;  
 Query Match 100.0%; Score 100; DB 1; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-08;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VERVDFTNHLEDFRRNINK 19  
 |||  
 Db 126 VERVDFTNHLEDFRRNINK 144

RESULT 2  
 SCCL\_HUMAN  
 ID SCCL\_HUMAN STANDARD; PRT; 390 AA.  
 AC P29508: 096J21;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Squamous cell carcinoma antigen 1 (SCCA-1) (Protein T4-A).  
 GN SERPINB3 OR SCCA1 OR SCCA.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE=92068241; PubMed=1958219;  
 RA Sumnaml Y., Kishi F., Sekiguchi K., Kato H.;  
 RT "Squamous cell carcinoma antigen is a new member of the serine  
 protease inhibitors";  
 RL Biochem. Biophys. Res. Commun. 181:51-58(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95241462; PubMed=7724531;  
 RA Schneider S.S., Schick C., Fish K.E., Miller E., Pena J.C.,  
 RA Treter S.D., Hul S.M., Silverman G.A.;  
 RT "A serine proteinase inhibitor locus at 18q21.3 contains a tandem  
 duplication of the human squamous cell carcinoma antigen gene";  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:3147-3151(1995).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lung;  
 RA Strassberg R.;  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: MAY ACT AS A PROTEASE INHIBITOR TO MODULATE THE HOST  
 IMMUNE RESPONSE AGAINST TUMOR CELLS.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- TISSUE SPECIFICITY: SQUAMOUS CELLS.  
 CC -1- DEVELOPMENTAL STAGE: ITS EXPRESSION IS CLOSELY RELATED TO CELLULAR  
 DIFFERENTIATION IN BOTH NORMAL AND MALIGNANT SQUAMOUS CELLS.  
 CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; S66896; AAB20405.1; -  
 CC EMBL; U19556; AAA97552.1; -  
 CC EMBL; U19568; AAA86317.1; -  
 CC EMBL; U19559; AAA86317.1; JOINED.  
 CC EMBL; U19560; AAA86317.1; JOINED.  
 CC EMBL; U19562; AAA86317.1; JOINED.  
 CC EMBL; U19565; AAA86317.1; JOINED.  
 CC EMBL; U19567; AAA86317.1; JOINED.  
 CC EMBL; U19562; AAA86316.1; -  
 CC EMBL; U19559; AAA86316.1; JOINED.  
 CC EMBL; U19560; AAA86316.1; JOINED.

DR EMBL; BC005224; AAH05224.1; -  
 DR PIR; JTO966; JTO966.  
 DR HSSP; P01008; IATH.  
 DR Genew; HGNC:10569; SERPINB3.  
 DR MIM; 600517; -  
 DR InterPro; IPR000215; Serpin.  
 DR Pfam; PF00079; serpin.1.  
 DR SMART; SM00093; SERPIN\_1.  
 DR PROSITE; PS00284; SERPIN\_1.  
 KW Serpin; Serine protease inhibitor.  
 FT ACT\_SITE 354 355 REACTIVE BOND.  
 FT CONFLICT 357 357 T -> A (IN REF. 1).  
 SQ SEQUENCE 390 AA; 44564 MW; E5F27F986C752CFA CRC64;  
 Query Match 50.0%; Score 50; DB 1; Length 390;  
 Best Local Similarity 55.6%; Pred. No. 1.2;  
 Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 1 VERVDFTNHLEDFRRNIN 18  
 |||  
 Db 131 VESVDFANADESRKRKIN 148

RESULT 3  
 EX5B\_BORBU  
 ID EX5B\_BORBU STANDARD; PRT; 1169 AA.  
 AC O51578;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Exodeoxyribonuclease V beta chain (EC 3.1.11.5).  
 GN RCB OR BB0633.  
 GN Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.  
 OC Borrelia burgdorferi (Lyme disease spirochete).  
 OX NCBI\_TaxID=139;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 35210 / B31;  
 RX MEDLINE=98065943; PubMed=9403685;  
 RA Fraser C.M., Casjens S., Huang W.M., Clayton R.A.,  
 RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,  
 RA Dougherty B., Tomb J.F., Fleischmann R.D., Richardson D.,  
 RA Peterson J., Kertlavage A.R., Quackenbush J., Salzberg S., Hanson M.,  
 RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,  
 RA Ullrich S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,  
 RA Smith H.O., Venter J.C.;  
 RT "Genomic sequence of a Lyme disease spirochete, Borrelia  
 burgdorferi";  
 RL Nature 390:580-586(1997).  
 CC -1- FUNCTION: REQUIRED FOR EFFICIENT DNA REPAIR; IT CATALYZES THE  
 UNWINDING OF DOUBLE-STRANDED DNA AND THE CLEAVAGE OF SINGLE-  
 STRANDED DNA AND IT STIMULATES LOCAL GENOMIC RECOMBINATION.  
 CC ALL OF THESE ACTIVITIES REQUIRE CONCOMITANT HYDROLYSIS OF ATP  
 (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: Exonucleolytic cleavage (in the presence of  
 CC ATP) in either 5'- to 3'- or 3'- to 5'-direction to yield 5'-  
 CC phosphooligonucleotides.  
 CC -1- SUBUNIT: CONSIST OF THREE SUBUNITS. RCB, RECC AND RECD  
 CC (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE HELICASE FAMILY. UVRD SUBFAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AE001164; AAC66981.1; -  
 CC HSSP; P56255; 1PJR.  
 CC TIGR; BB0633; -

```

DR InterPro: IPR004586; RecB.
DR InterPro: IPR0000212; UvrD-helicase.
DR Pfam: PF00580; UvrD-helicase; 1.
DR TIGRFAMs: TIGR00609; recB; 1.
RM HydroLase; Nuclease; Endonuclease; Helicase; ATP-binding;
RV DNA repair; Complete proteome.
FW NP_BIND
FT 18
SO SEQUENCE 1169 AA; 137828 MW; B61D63C1C559B91F CRC64;

Query Match      49.0%; Score 49; DB 1; Length 1169;
Best Local Similarity 42.1%; Pred. No. 5.8;
Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

OY 1 VERVDFTNHLEDTRRNINK 19
    ::::|||||::: 11
Db 459 IEKIDFTNLSLPQKNNNKK 477

RESULT 4
NC04_HUMAN STANDARD; PRT; 210 AA.
ID NC04_HUMAN
AC Q43402;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Neighbor of COX4.
GN NC04 OR COX4L.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
CC NCBI_Taxid=9606;
LN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Retina, and Placenta;
RX MEDLINE=99269920; PubMed=10337626;
RA Bachman N.J., Wu W., Schmidt T.R., Grossman L.I., Lomax M.I.;
RT "The 5-prime region of the COX4 gene contains a novel overlapping
RL gene, NC04."
RL Mamm. Genome 10:506-512(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-Placenta;
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- TISSUE SPECIFICITY: Expressed in liver, pancreas, heart, lung,
CC kidney, brain, skeletal muscle, and placenta. Expression levels
CC are highest in pancreas and moderate in heart, skeletal muscle,
CC and placenta.
CC -1- SIMILARITY: BELONGS TO THE UPF0172 (NC04) FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/isb-ch).
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: AF005888; AAB94489.1; -
DR EMBL: AF005889; AAB94489.1; -
DR EMBL: BC005886; AAOH05886.1; -
DR EMBL: BC001472; AAH01472.1; -
DR Genew; HGNC:7864; NC04.
DR MIM: 604886; -
DR InterPro: IPR005366; UPF0172.
DR Pfam: PF03665; UPF0172.1.
SO SEQUENCE 210 AA; 23773 MW; 41367FDD98769250 CRC64;

Query Match      48.5%; Score 48.5; DB 1; Length 210;
Best Local Similarity 52.4%; Pred. No. 1;
Matches 11; Conservative 2; Mismatches 3; Indels 5; Gaps 1;

OY 4 VDFTNHLEDTRRN----INK 19

```

```

Db      184 VDFPNHLDLRDWTNPBINK 204
      ||| |||:|: |||
      184 VDFPNHLDLRDWTNPBINK 204

RESULT 5
NOC4_MOUSE
ID NOC4_MOUSE STANDARD: PRT: 207 AA.
AC 070378:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Neighbor of COX4.
GN NOC4 OR COX4AL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Kidney;
RX MEDLINE=99269920; PubMed=10337626;
RA Bachman N.J., Wu W., Schmidt T.R., Grossman L.I., Lomax M.I.;
RT "The 5'-prime region of the COX4 gene contains a novel overlapping
   gene, NOC4."
RL Mamm. Genome 10:506-512(1999).
CC -
CC -I- SIMILARITY: BELONGS TO THE UPP0172 (NOC4) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF052621; AAC12933.1; -
DR MGD: MGI:1343095; NOC4.
DR InterPro: IPR005366; UPP0172.
DR Pfam: PF03665; UPP0172; 1.
SQ SEQUENCE 207 AA; 23348 MW; 8F3CEP1BF2498958 CRC64;

Query Match 47.5%; Score 47.5; DB 1; Length 207;
Best Local Similarity 52.4%; Pred. No. 1.5;
Matches 11; Conservative 2; Mismatches 3; Indels 5; Gaps 1;

OY 4 VDFTNHLEDTRRN-----INK 19
      ||| |||:|: |||
Db      181 VDFPNHLDLRDWTNPBINK 201

RESULT 6
CP7A_RABIT
ID CP7A_RABIT STANDARD: PRT: 501 AA.
AC P51542:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Cytochrome P450 7A1 (Cholesterol 7-alpha-monooxygenase) (CYP7A1)
DE (EC 1.14.13.17) (Cholesterol 7-alpha-hydroxylase).
GN CYP7A1 OR CYP7.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxId=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=New Zealand white; TISSUE=Liver;
RX MEDLINE=95271192; PubMed=7751825;
RA Kai M., Eto T., Kondo K., Setoguchi Y., Higashi S., Maeda Y.,
RA Setoguchi T.;
RT "Synchronous circadian rhythms of mRNA levels and activities of
   cholesterol 7-alpha-hydroxylase in the rabbit and rat.";
RL J. Lipid Res 36:367-374(1995).

```

```

CC -1- CATALYTIC ACTIVITY: Cholesterol + NADPH + O(2) = 7-alpha-
CC Hydroxysterol + NADP(+) + H(2)O.
CC -1- PATHWAY: FIRST AND RATE-LIMITING STEP IN THE CONVERSION OF
CC CHOLESTEROL TO BILE ACIDS.
CC -1- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L10754; AAA74382.1; -
DR InterPro: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; P450: 1
DR PROSITE, PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW Microsome; Endoplasmic reticulum; Cholesterol metabolism; NADP.
FT BINDING 441 441 HEME (BY SIMILARITY).
SQ SEQUENCE 501 AA; 58090 MW; FEF1247B151AA6B1 CRC64;

Query Match 47.0%; Score 47; DB 1; Length 501;
Best Local Similarity 66.7%; Pred. No. 4.7;
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 5 DFTNHEDETRRNINR 19
Db 366 DFTLHEDGSYNIRK 380

RESULT 7
CP7A_MOUSE STANDARD; PRT; 503 AA.
AC Q64505;
DT 13-JUL-1999 (Rel. 38, Created)
DT 13-JUL-1999 (Rel. 38, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Cytochrome P450 7A1 (Cholesterol 7-alpha-monooxygenase) (CYPVII)
DE (EC 1.14.13.17) (Cholesterol 7-alpha-hydroxylase).
GN CYP7A1 OR CYP7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=94375022; PubMed=8088795;
RA Tzung K.W., Ishimura-Oka K., Kihara S., Oka K., Chan L.;
RT "Structure of the mouse cholesterol 7 alpha-hydroxylase gene.";
RT Genomics 21:244-247(1994).
CC -1- CATALYTIC ACTIVITY: Cholesterol + NADPH + O(2) = 7-alpha-
CC hydroxysterol + NADP(+) + H(2)O.
CC -1- PATHWAY: FIRST AND RATE-LIMITING STEP IN THE CONVERSION OF
CC CHOLESTEROL TO BILE ACIDS.
CC -1- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L23754; AAA68867.1; -
DR MGD: MGI:106091; CYP7A1.
DR InterPro: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; P450: 1.
DR PRINTS: PR00385; P450.

```

```

DR PROSITE, PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW Microsome; Endoplasmic reticulum; Cholesterol metabolism; NADP.
FT BINDING 444 444 HEME (BY SIMILARITY).
SQ SEQUENCE 503 AA; 57258 MW; 51DB3C71D6C79791 CRC64;

Query Match 47.0%; Score 47; DB 1; Length 503;
Best Local Similarity 66.7%; Pred. No. 4.8;
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 5 DFTNHEDETRRNINR 19
Db 369 DFTLHEDGSYNIRK 383

RESULT 8
CP7A_RAT STANDARD; PRT; 503 AA.
AC P18125; P51543;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cytochrome P450 7A1 (Cholesterol 7-alpha-monooxygenase) (CYPVII)
DE (EC 1.14.13.17) (Cholesterol 7-alpha-hydroxylase).
GN CYP7A1 OR CYP7.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Sprague-Dawley; TISSUE=Liver;
RC MEDLINE=90307735; PubMed=1694852;
RX Li Y.C., Wang D.P., Chiang J.Y.L.;
RA "Regulation of cholesterol 7 alpha-hydroxylase in the liver. Cloning,
RT sequencing, and regulation of cholesterol 7 alpha-hydroxylase mRNA.";
RT J. Biol. Chem. 265:12012-12019(1990).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=Sprague-Dawley;
RC MEDLINE=90243699; PubMed=2335522;
RX Uehinek D.F., Andersson S., Slaughter C.A., Russell D.W.;
RA "Cloning and regulation of cholesterol 7 alpha-hydroxylase, the rate-
RT limiting enzyme in bile acid biosynthesis.";
RT J. Biol. Chem. 265:8190-8197(1990).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=90033362; PubMed=2806567;
RA Noshiro M., Nishimoto M., Morohashi K., Okuda K.;
RT "Molecular cloning of cDNA for cholesterol 7 alpha-hydroxylase from
RT rat liver microsomes. Nucleotide sequence and expression.";
RT FEBS Lett. 257:97-100(1989).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=90277612; PubMed=1693613;
RA Noshiro M., Nishimoto M., Okuda K.;
RT "Rat liver cholesterol 7 alpha-hydroxylase. Pretranslational
RT regulation for circadian rhythm.";
RT J. Biol. Chem. 265:10036-10041(1990).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=91177904; PubMed=2007596;
RA Nishimoto M., Gotoh O., Okuda K., Noshiro M.;
RT "Structural analysis of the gene encoding rat cholesterol alpha-
RT hydroxylase, the key enzyme for bile acid biosynthesis.";
RT J. Biol. Chem. 266:6467-6471(1991).
RN [6]
RP SEQUENCE OF 1-325 FROM N.A.
RA STRAIN=Sprague-Dawley;
RC MEDLINE=94292507; PubMed=8021257;
RX Chiang J.Y., Stroup D.;
RT "Identification and characterization of a putative bile acid-

```

```

RT responsive element in cholesterol 7 alpha-hydroxylase gene promoter."
RL J. Biol. Chem. 269:17502-17507(1994).
RN [7]
RP SEQUENCE OF 1-27 FROM N.A.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=91084435; PubMed=2261433;
RA Jellinek D.F., Russell D.W.;
RT "Structure of the rat gene encoding cholesterol 7 alpha-hydroxylase.";
RL Biochemistry 29:7781-7785(1990).
RN [8]
RP SEQUENCE OF 1-26 FROM N.A.
RX MEDLINE=93041942; PubMed=1420318;
RA Chiang J.Y., Yang T.P., Mang D.P.;
RT "Cloning and 5'-flanking sequence of a rat cholesterol 7 alpha-
hydroxylase gene.";
RL Biochem. Biophys. Acta 1132:337-339(1992).
CC -1 CATALYTIC ACTIVITY: Cholesterol + NADPH + O(2) -> 7-alpha-
hydroxycholesterol + NADP(+).
CC -1 ENZYME REGULATION: INHIBITION OF ACTIVITY BY BILE ACID FEEDBACK.
CC -1 PATHWAY: FIRST AND RATE-LIMITING STEP IN THE CONVERSION OF
CHOLESTEROL TO BILE ACIDS.
CC -1 SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC -1 TISSUE SPECIFICITY: Liver.
CC -1 INDUCTION: BY CHOLESTYRAMINE AND CHOLESTEROL.
CC -1 SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC -----
DR EMBL: J05509; AAA40839.1; -
DR EMBL: J02926; AAA40923.1; -
DR EMBL: J05430; AAA41041.1; -
DR EMBL: J05460; AAA03649.1; -
DR EMBL: M59189; AAA41042.1; -
DR EMBL: M59184; AAA41042.1; JOINED.
DR EMBL: M59185; AAA41042.1; JOINED.
DR EMBL: M59186; AAA41042.1; JOINED.
DR EMBL: M59187; AAA41042.1; JOINED.
DR EMBL: M59188; AAA41042.1; JOINED.
DR EMBL: U01962; AAA21144.2; -
DR EMBL: X17993; CAB57878.1; -
DR EMBL: Z14108; CAB78481.1; -
DR PIR: A35376; A35376.
DR InterPro: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; p450.1
DR PROSITE: PS00086; CYTOCHROME_P450.1.
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KM Microsome; Endoplasmic reticulum; Cholesterol metabolism; NADP.
FT BINDING 444 HEME (BY SIMILARITY).
FT CONFLICT 371 T -> S (IN REF. 5).
SQ SEQUENCE 503 AA; 56882 MW; EA825AA6E74BF5F6 CRC64;
Query Match 47.0%; Score 47; DB 1; Length 503;
Best Local Similarity 66.7%; Pred. No. 4.8;
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

DE Cytochrome P450 7A1 (Cholesterol 7-alpha-monooxygenase) (CYP7A1)
DE (EC 1.14.13.17) (Cholesterol 7-alpha-hydroxylase).
GN CYP7A1 OR CYP7.
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Cricetinae;
OC Cricetulus.
OX NCBI_TaxId=10029;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=94029018; PubMed=8105753;
RA Crestani M., Galli G., Chiang J.Y.;
RT "Genomic cloning, sequencing, and analysis of the hamster cholesterol
7 alpha-hydroxylase gene (CYP7).";
RL Arch. Biochem. Biophys. 306:451-460(1993).
CC -1 CATALYTIC ACTIVITY: Cholesterol + NADPH + O(2) -> 7-alpha-
hydroxycholesterol + NADP(+).
CC -1 PATHWAY: FIRST AND RATE-LIMITING STEP IN THE CONVERSION OF
CHOLESTEROL TO BILE ACIDS.
CC -1 SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC -1 SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC -----
DR EMBL: L04690; AAA03751.1; -
DR InterPro: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; p450.1.
DR PROSITE: PS00086; CYTOCHROME_P450.1.
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KM Microsome; Endoplasmic reticulum; Cholesterol metabolism; NADP.
FT BINDING 444 HEME (BY SIMILARITY).
SQ SEQUENCE 504 AA; 57447 MW; BDSE521D0C195257 CRC64;
Query Match 47.0%; Score 47; DB 1; Length 504;
Best Local Similarity 66.7%; Pred. No. 4.8;
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

DB 5 DFTNLEDRNRINK 19
   ||| ||| ||| |||
369 DFTLHEDGYNIRK 383

```

```

RESULT 10
CP7A_HUMAN
ID CP7A_HUMAN STANDARD; PRT; 504 AA.
AC P22680;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Cytochrome P450 7A1 (Cholesterol 7-alpha-monooxygenase) (CYP7A1)
DE (EC 1.14.13.17) (Cholesterol 7-alpha-hydroxylase).
GN CYP7A1 OR CYP7.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93176797; PubMed=8439551;
RA Nishimoto M., Noshiro M., Okuda K.;
RT "Structure of the gene encoding human liver cholesterol 7 alpha-
hydroxylase.";
RL Biochim. Biophys. Acta 1172:147-150(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90346120; PubMed=2384150;

```

RA Noshiro M., Okuda K.;  
 RT "Molecular cloning and sequence analysis of cDNA encoding human  
 RL cholesterol 7 alpha-hydroxylase.";  
 RN FEBS Lett. 268:137-140(1990).  
 [3]  
 RX MEDLINE=92304280; PubMed=1610352;  
 RP MEDLINE=92304280; PubMed=1610352;  
 RA Karam W.G., Chiang J.Y.;  
 RT "Polymorphisms of human cholesterol 7 alpha-hydroxylase.";  
 RL Biochem. Biophys. Res. Commun. 185:588-595(1992).  
 CC -1- CATALYTIC ACTIVITY: Cholesterol + NADPH + O(2) = 7-alpha-  
 CC hydroxycholesterol + NADP(+) + H(2)O.  
 CC -1- PATHWAY: FIRST AND RATE-LIMITING STEP IN THE CONVERSION OF  
 CC CHOLESTEROL TO BILE ACIDS.  
 CC -1- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X56088; CA39568.1; -;  
 DR EMBL: M93133; AA58435.1; -;  
 DR PIR: JH0659; JH0659.  
 DR PIR: S29818; S29818.  
 DR GeneW: HGNC:2651; CYP7A1.  
 DR MIM: 118455; -;  
 DR InterPro: IPR001128; Cytochrome\_P450.  
 DR Pfam: PF00067; P450; 1.  
 DR PROSITE: PS00086; CYTOCHROME\_P450; 1.  
 KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;  
 KW Microsome; Endoplasmic reticulum; Cholesterol metabolism; NADP;  
 KW Polymorphism.  
 FT BINDING 444 444 HEME (BY SIMILARITY).  
 FT VARIANT 100 100 F -> S.  
 FT FTID-VAR\_001259.  
 FT CONFLICT 347 347 D -> N (IN REF. 2).  
 FT CONFLICT 385 385 D -> S (IN REF. 2).  
 SQ SEQUENCE 504 AA: 57660 MW: D8067E0FF6342949 CRC64;  
 Query Match 47.0%; Score 47; DB 1; Length 504;  
 Best Local Similarity 66.7%; Pred. No. 4.8;  
 Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 5 DFTNHLEDTRRNRNK 19  
 DB 369 DFTLHLEDSYNRK 383  
 RESULT 11  
 MASP\_HUMAN  
 ID MASP\_HUMAN STANDARD: PRT; 375 AA.  
 AC P36952;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Maspilin precursor (protease inhibitor 5).  
 GN SERPINB5 OR P15.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Mammary gland;  
 RX MEDLINE=94120413; PubMed=8290962;  
 RA Zou Z., Antisiewicz A., Hendrix M.J.C., Thor A., Neveu M., Sheng S.,  
 RA Rafidi K., Sedor E., Sager R.;  
 RT "Maspilin, a serpin with tumor-suppressing activity in human mammary

RT epithelial cells.";  
 RL Science 263:526-529(1994).  
 RN [2]  
 RX MEDLINE=95318177; PubMed=7797587;  
 RP MEDLINE=95318177; PubMed=7797587;  
 RA Pemberton P.A., Wong D.T., Gibson H.L., Klefer M.C.,  
 RA Fitzpatrick P.A., Sager R., Barr P.J.;  
 RT "The tumor suppressor maspin does not undergo the stressed to relaxed  
 RT transition or inhibit trypsin-like serine proteases. Evidence that  
 RT maspin is not a protease inhibitory serpin.";  
 RL J. Biol. Chem. 270:15832-15837(1995).  
 CC -1- FUNCTION: TUMOR SUPPRESSOR. IT BLOCKS THE GROWTH, INVASION, AND  
 CC METASTATIC PROPERTIES OF MAMMARY TUMORS. AS IT DOES NOT UNDERGO  
 CC THE S (STRESSED) TO R (RELAXED) CONFORMATIONAL TRANSITION  
 CC CHARACTERISTIC OF ACTIVE SERPINS, IT EXHIBITS NO SERINE PROTEASE  
 CC INHIBITORY ACTIVITY.  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- TISSUE SPECIFICITY: NORMAL MAMMARY EPITHELIAL CELLS.  
 CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U04313; AA18957.1; -;  
 DR HSSP: P05619; 1HLE.  
 DR GeneW: HGNC:8949; SERPINB5.  
 DR MIM: 154790; -;  
 DR InterPro: IPR000240; Maspilin.  
 DR InterPro: IPR000215; Serpin.  
 DR Pfam: PF00079; serpin; 1.  
 DR PRINTS: PR000676; MASPILN.  
 DR SMART: SM00093; SERPIN; 1.  
 DR PROSITE: PS00284; SERPIN; 1.  
 KW Serpin; Glycoprotein; Signal.  
 FT SIGNAL 1 2  
 FT CHAIN 1 375  
 FT ACT\_SITE 340 341 REACTIVE BOND (BY SIMILARITY).  
 FT CARBOHYD 99 99 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 361 361 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 375 AA: 42138 MW: F2D49FDCD6AC3DC2 CRC64;  
 Query Match 46.0%; Score 46; DB 1; Length 375;  
 Best Local Similarity 50.0%; Pred. No. 5;  
 Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 VERVDFTNHLEDTRRNRIN 18  
 DB 116 LETVDKDKLEETKQGIN 133  
 RESULT 12  
 MASP\_MOUSE  
 ID MASP\_MOUSE STANDARD: PRT; 375 AA.  
 AC F70124;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Maspilin precursor (protease inhibitor 5).  
 GN SERPINB5 OR P15 OR SP17.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97206139; PubMed=9132279;

RA Zhang M., Sheng S., Maass N., Sager R.;  
 RT "maspin: the mouse homolog of a human tumor suppressor gene inhibits  
 RT mammary tumor invasion and motility."; *Mol. Med.* 3:49-59(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Strauberg R.;  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: TUMOR SUPPRESSOR. IT BLOCKS THE GROWTH, INVASION, AND  
 CC METASTATIC PROPERTIES OF MAMMARY TUMORS. AS IT DOES NOT UNDERGO  
 CC THE S (STRESSED) TO R (RELAXED) CONFORMATIONAL TRANSITION  
 CC CHARACTERISTIC OF ACTIVE SERPINS, IT EXHIBITS NO SERINE PROTEASE  
 CC INHIBITORY ACTIVITY (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U54705; AAB06042.1; -  
 DR EMBL: BC005434; AA05434.1; -  
 DR HSSP: P05619; 1HLE.  
 DR MGI: 109579; Serpinb5.  
 DR InterPro: IPR000240; Maspin.  
 DR InterPro: IPR000215; Serpin.  
 DR Pfam: PF00079; serpin; 1.  
 DR PRINTS: PRO0676; MASPIN.  
 DR SMART: SM00093; SERPIN; 1.  
 DR PROSITE: PS00284; SERPIN; 1.  
 DR Serpin; Glycoprotein; Signal.  
 KW CHAIN 1  
 FT SIGNAL 1  
 FT ACT\_SITE 340 341 REACTIVE BOND (BY SIMILARITY).  
 FT CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 298 298 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 361 361 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SO SEQUENCE 375 AA; 42111 MW; 5894DD8B9A76F9A7 CRC64;  
 Query Match 46.0%; Score 46; DB 1; Length 375;  
 Best Local Similarity 50.0%; Pred. No. 5;  
 Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 VERVDFTHNLEDTRRNIN 18  
 DB 116 LETVDFKDLKETKQIN 133  
 RESULT 13  
 MASP\_RAT  
 ID MASP\_RAT STANDARD; PRT; 375 AA.  
 AC -P70564;  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Maspin precursor (Protease inhibitor 5).  
 GN SERPINB5 OR P15.  
 OS Rattus norvegicus (Rat).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Sprague-Dawley; TISSUE-Vagina;  
 RX MEDLINE=97218145; PubMed=9065806;  
 RA Umekita Y., Hikapka R.A., Lleo S.;  
 RT "Rat and human maspins: structures, metastatic suppressor activity  
 RT and mutation in prostate cancer cells."; *Cancer Lett.* 113:87-93(1997).

CC -1- FUNCTION: TUMOR SUPPRESSOR. IT BLOCKS THE GROWTH, INVASION, AND  
 CC METASTATIC PROPERTIES OF MAMMARY TUMORS. AS IT DOES NOT UNDERGO  
 CC THE S (STRESSED) TO R (RELAXED) CONFORMATIONAL TRANSITION  
 CC CHARACTERISTIC OF ACTIVE SERPINS, IT EXHIBITS NO SERINE PROTEASE  
 CC INHIBITORY ACTIVITY (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U58657; AAB06043.1; -  
 DR HSSP: P05619; 1HLE.  
 DR InterPro: IPR000240; Maspin.  
 DR InterPro: IPR000215; Serpin.  
 DR Pfam: PF00079; serpin; 1.  
 DR PRINTS: PRO0676; MASPIN.  
 DR SMART: SM00093; SERPIN; 1.  
 DR PROSITE: PS00284; SERPIN; 1.  
 DR Serpin; Glycoprotein; Signal.  
 KW CHAIN 1  
 FT SIGNAL 1  
 FT ACT\_SITE 340 341 REACTIVE BOND (BY SIMILARITY).  
 FT CARBOHYD 99 99 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 298 298 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 361 361 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SO SEQUENCE 375 AA; 42063 MW; E82C4E6A0F22482 CRC64;  
 Query Match 46.0%; Score 46; DB 1; Length 375;  
 Best Local Similarity 50.0%; Pred. No. 5;  
 Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 VERVDFTHNLEDTRRNIN 18  
 DB 116 LETVDFKDLKETKQIN 133  
 RESULT 14  
 SCC2\_HUMAN  
 ID SCC2\_HUMAN STANDARD; PRT; 390 AA.  
 AC P48594;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Squamous cell carcinoma antigen 2 (SCCA-2) (Leuplin).  
 GN SERPINB4 OR SCCA2.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95241462; PubMed=7724531;  
 RA Schneider S.S., Schick C., Fish K.E., Miller E., Pena J.C.,  
 RA Treter S.D., Hui S.M., Silverman G.A.;  
 RT "A serine proteinase inhibitor locus at 18q21.3 contains a tandem  
 RT duplication of the human squamous cell carcinoma antigen gene";  
 RT Proc. Natl. Acad. Sci. U.S.A. 92:3147-3151(1995).  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96013887; PubMed=7589435;  
 RA Barnes R.C., Worrall D.M.;  
 RT "Identification of a novel human serpin gene: cloning sequencing and  
 RT expression of leuplin."; *FEBS Lett.* 373:61-65(1995).  
 RL [3]

```

RP SEQUENCE FROM N.A.
RX MEDLINE-21167379; PubMed-11267667;
RA Hamada K., Shinomiya H., Asano Y., Kihana T., Iwamoto M., Hanakawa Y.,
RA Hashimoto K., Hirose S., Kyo S., Ito M.;
RT "Molecular cloning of human squamous cell carcinoma antigen 1 gene and
RT characterization of its promoter.";
RL Biochim. Biophys. Acta 1518:124-131(2001).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE-Bone marrow;
RA Strausberg R.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAY ACT AS A PROTEASE INHIBITOR TO MODULATE THE HOST
CC IMMUNE RESPONSE AGAINST TUMOR CELLS.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: SQUAMOUS CELLS.
CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U19576; AAA92602.1; -.
DR EMBL: U19570; AAA92602.1; JOINED.
DR EMBL: U19571; AAA92602.1; JOINED.
DR EMBL: U19572; AAA92602.1; JOINED.
DR EMBL: U19574; AAA92602.1; JOINED.
DR EMBL: U19575; AAA92602.1; JOINED.
DR EMBL: U19557; AAA97553.1; -.
DR EMBL: X89015; CAA61420.1; -.
DR EMBL: AB035089; BAB21525.1; -.
DR EMBL: BC017401; AAH17401.1; -.
DR HSSP: P01008; IATH.
DR Genew; HGNC:10570; SERPINB4.
DR MIM; 600518; -.
DR InterPro: IPR000215; Serpin.
DR Pfam: PF000079; serpin; 1.
DR SMART: SM00093; SERPIN_1.
DR PROSITE; PS00284; SERPIN; 1.
KW Serpin; Serine protease inhibitor.
FT ACT_SITE 354 355 REACTIVE BOND.
SQ SEQUENCE 390 AA; 44854 MW; 0AE213CD892587D5 CRC64;

Query Match 46.0%; Score 46; DB 1; Length 390;
Best Local Similarity 50.0%; Pred. No. 5.2;
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 VERVDFTNHLEDTNRNIN 18
Db 131 VESTDFANPEESRKIN 148

RESULT 15
BXG_CLOBO STANDARD; PRT; 1296 AA.
AC 060393;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type G precursor (EC 3.4.24.69) (BONT/G)
DE (Bontoxilysin G).
GN BONTG
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=113 / 30;

```

```

RX MEDLINE-94092745; PubMed-8268233;
RA Campbell K., Collins M.D., East A.K.;
RT "Nucleotide sequence of the gene coding for Clostridium botulinum
RT (Clostridium argentineense) type G neurotoxin: genealogical comparison
RT with other Clostridial neurotoxins.";
RL Biochim. Biophys. Acta 1216:487-491(1993).
RN [1]
RP SEQUENCE: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE.
CC -1- CATALYTIC ACTIVITY: limited hydrolysis of proteins of the
CC neuroexcitatory apparatus, synaptobrevins, SNAP25 or syntaxin. NO
CC detected action on small molecule substrates.
CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
CC WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
CC FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC -1- SUBCELLULAR LOCATION: secreted (by similarity).
CC -1- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X74162; CAA52275.1; -.
DR HSSP: P10845; 3BTA.
DR MEROPS; M27.002; -.
DR InterPro: IPR000395; Bontoxilysin.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOTOXILYSIN.
DR Prodom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Neurotoxin; Hydrolase; Metalloprotease; Zinc.
FT INIT_MET 0
FT CHAIN 1 441
FT CHAIN 442 1296
FT METAL 229 229
FT ACT_SITE 230 230
FT METAL 233 233
FT DISULFID 435 449
SQ SEQUENCE 1296 AA; 149013 MW; DC8E47E15F65C31 CRC64;

Query Match 46.0%; Score 46; DB 1; Length 1296;
Best Local Similarity 47.4%; Pred. No. 20;
Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 VERVDFTNHLEDTNRNIN 19
Db 554 IENVLQTLNSLDALRNK 572

```

Search completed: July 11, 2003, 11:55:14  
 Job time : 3.81705 secs